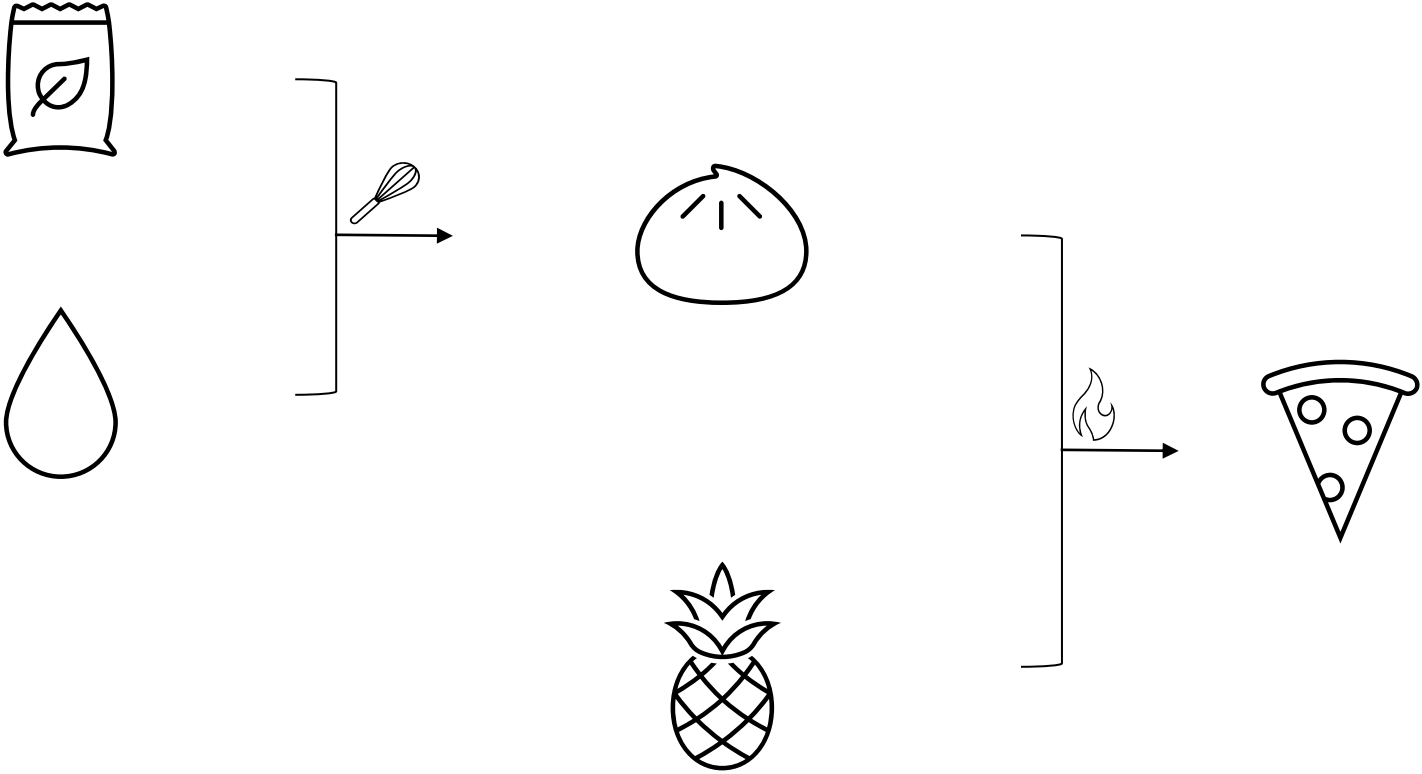


Fake it till you make it: planning chemical syntheses for drug discovery

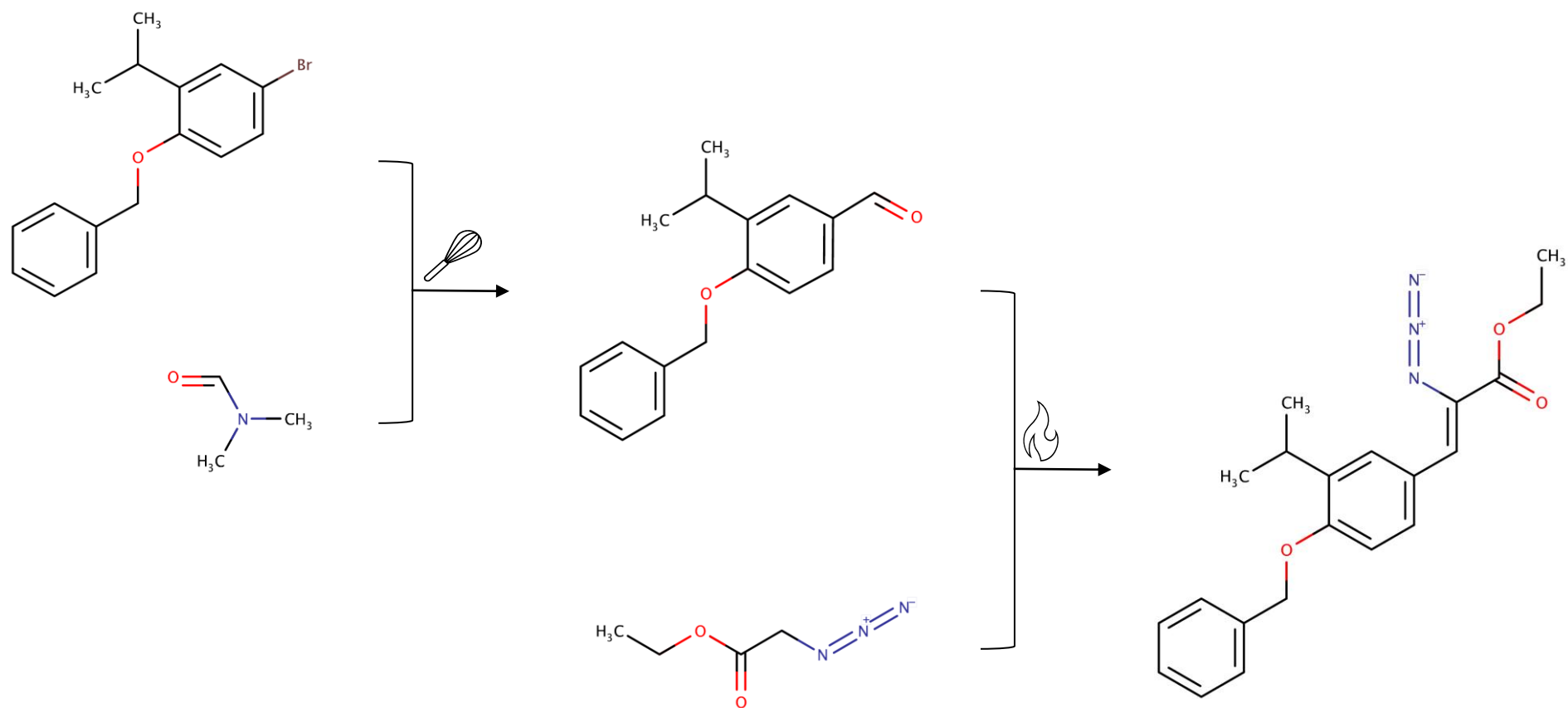
Krzysztof Maziarz
Microsoft Research AI for Science



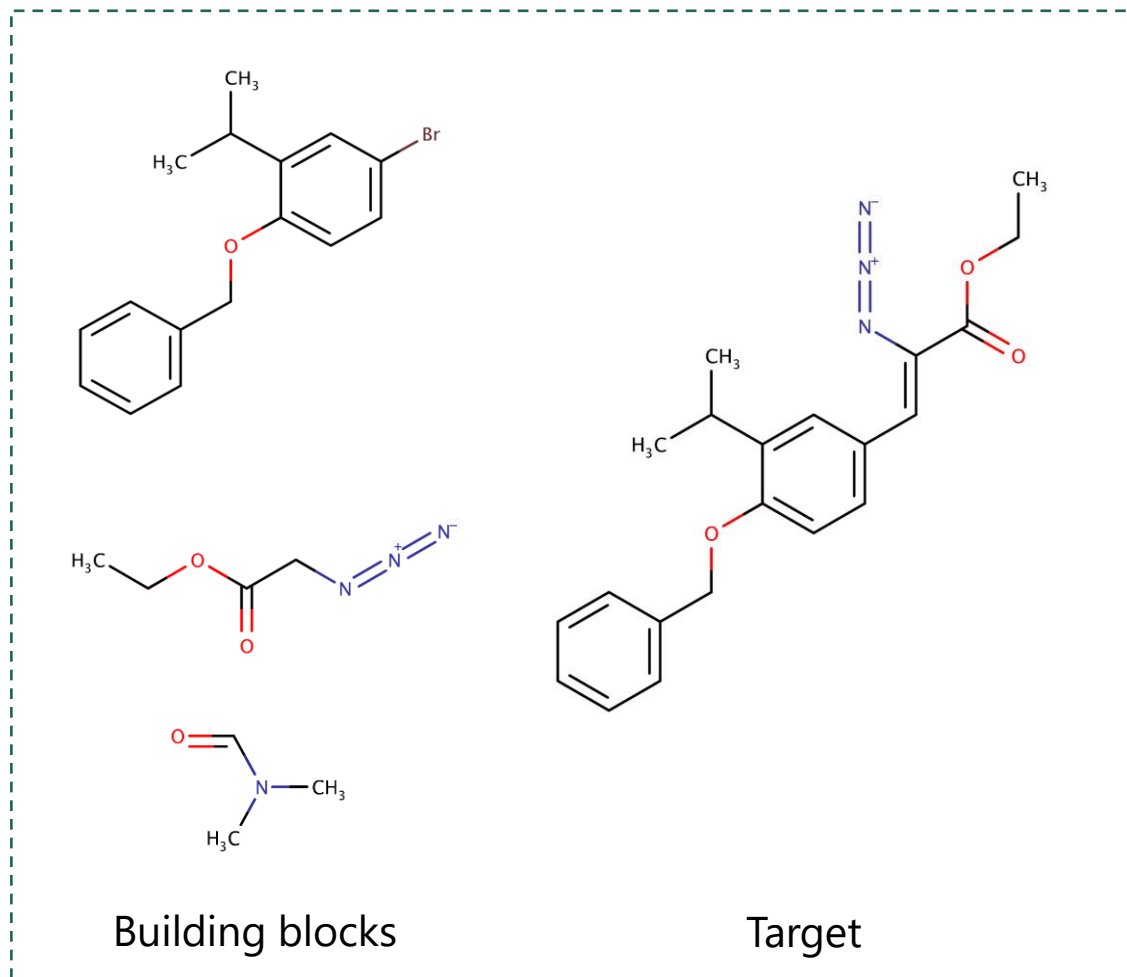
Synthesis: cooking with molecules



Synthesis: cooking with molecules

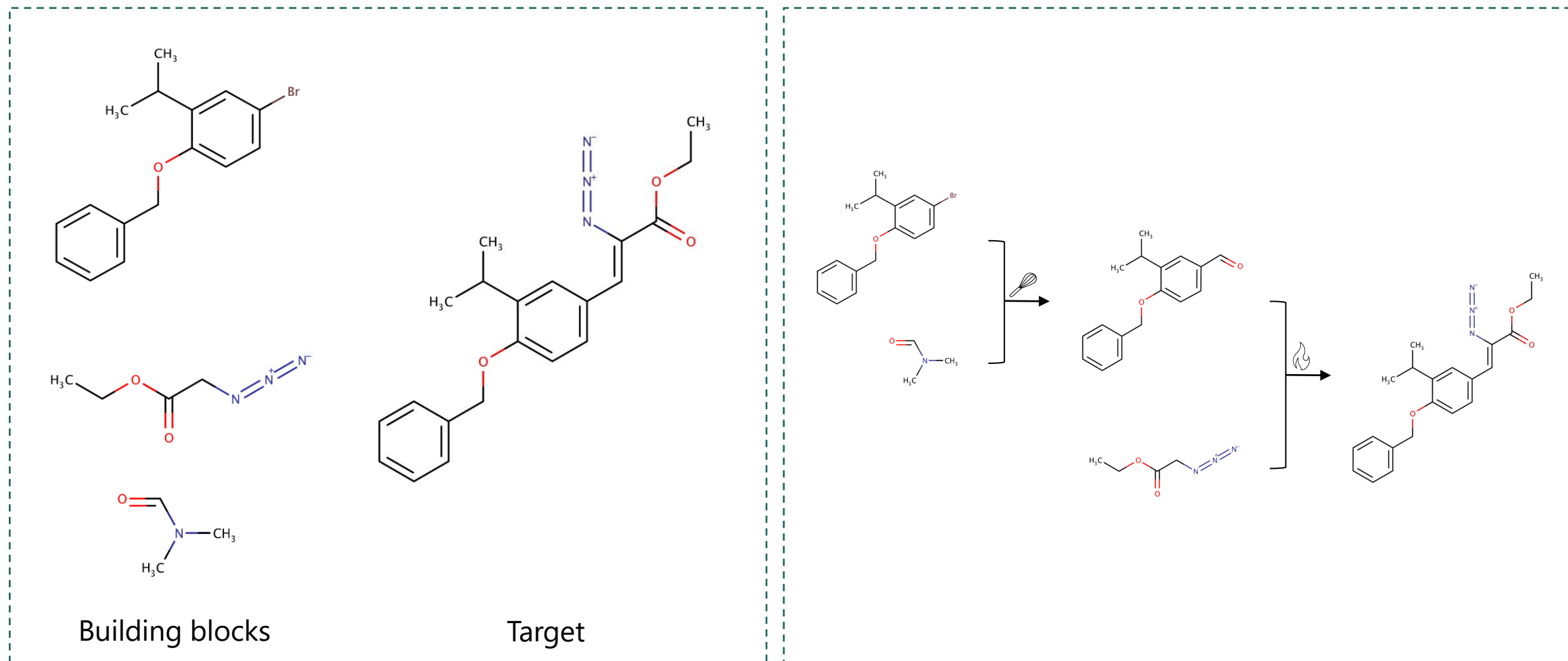


Synthesis planning



Input

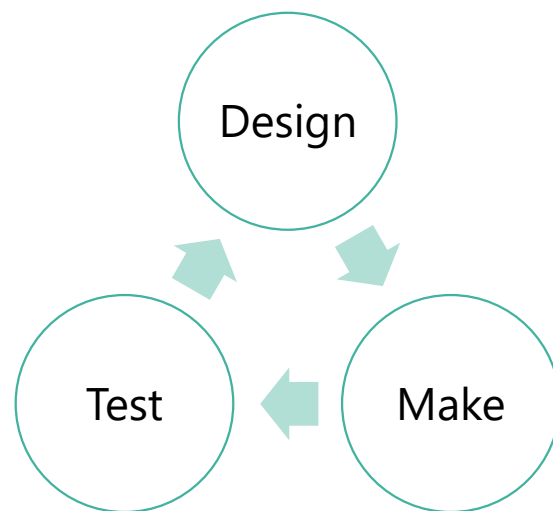
Synthesis planning



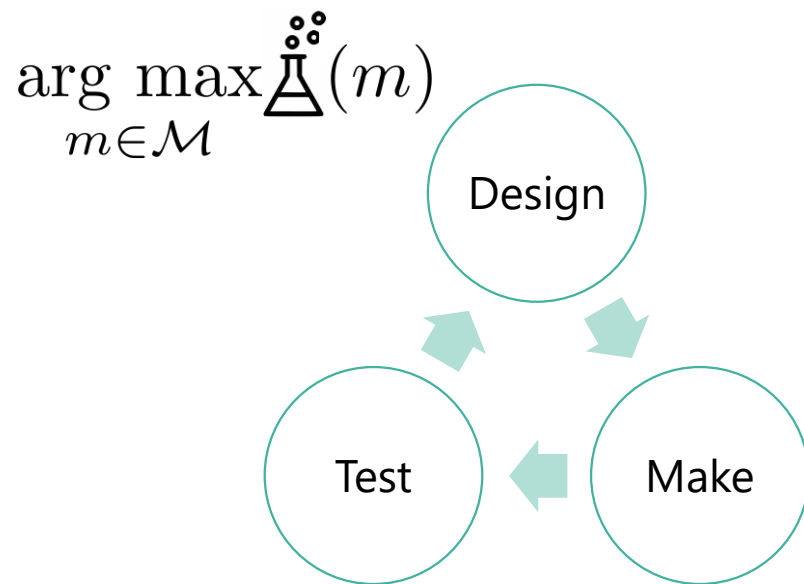
Input

Output (synthetic route)

Drug discovery, or why we need synthesis planning

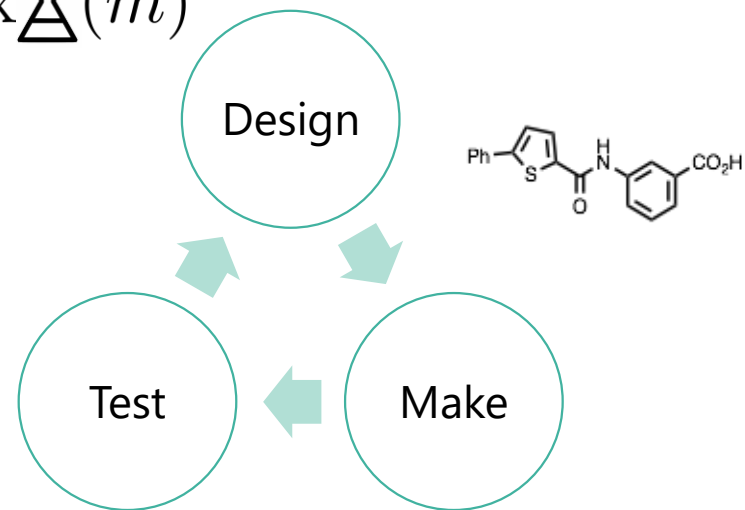


Drug discovery, or why we need synthesis planning

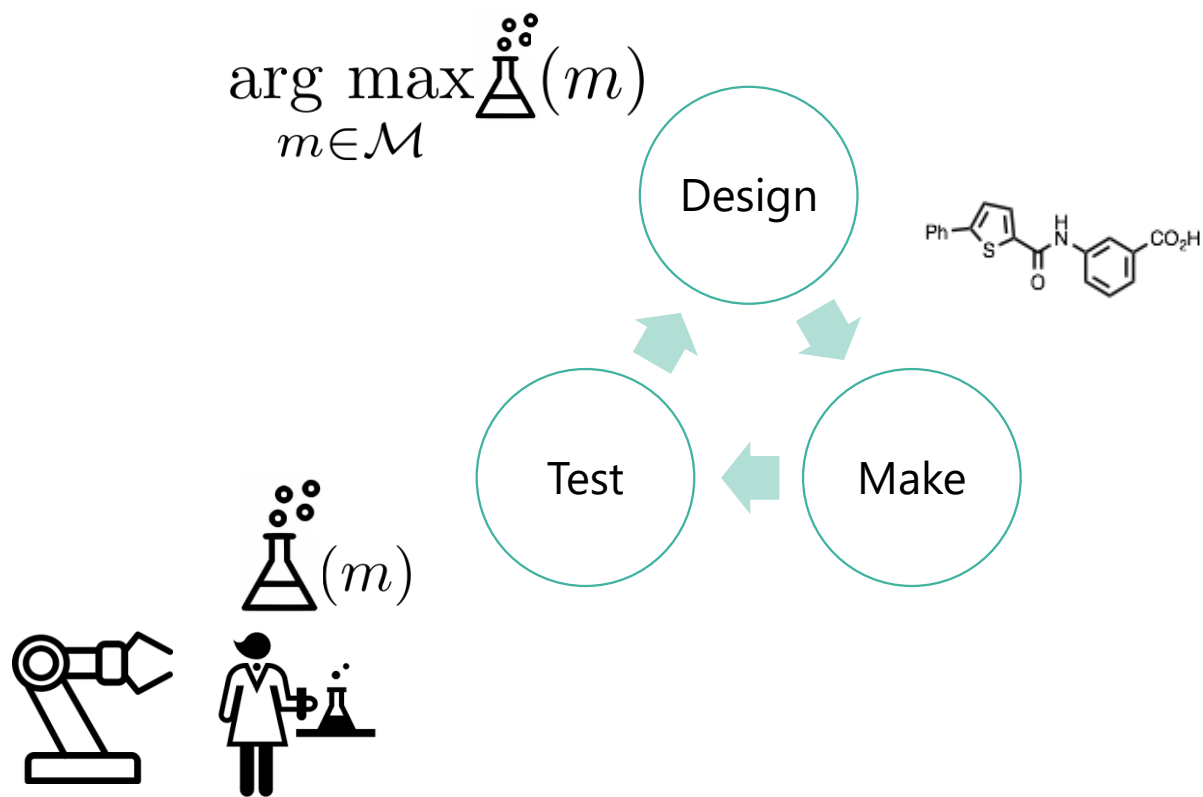


Drug discovery, or why we need synthesis planning

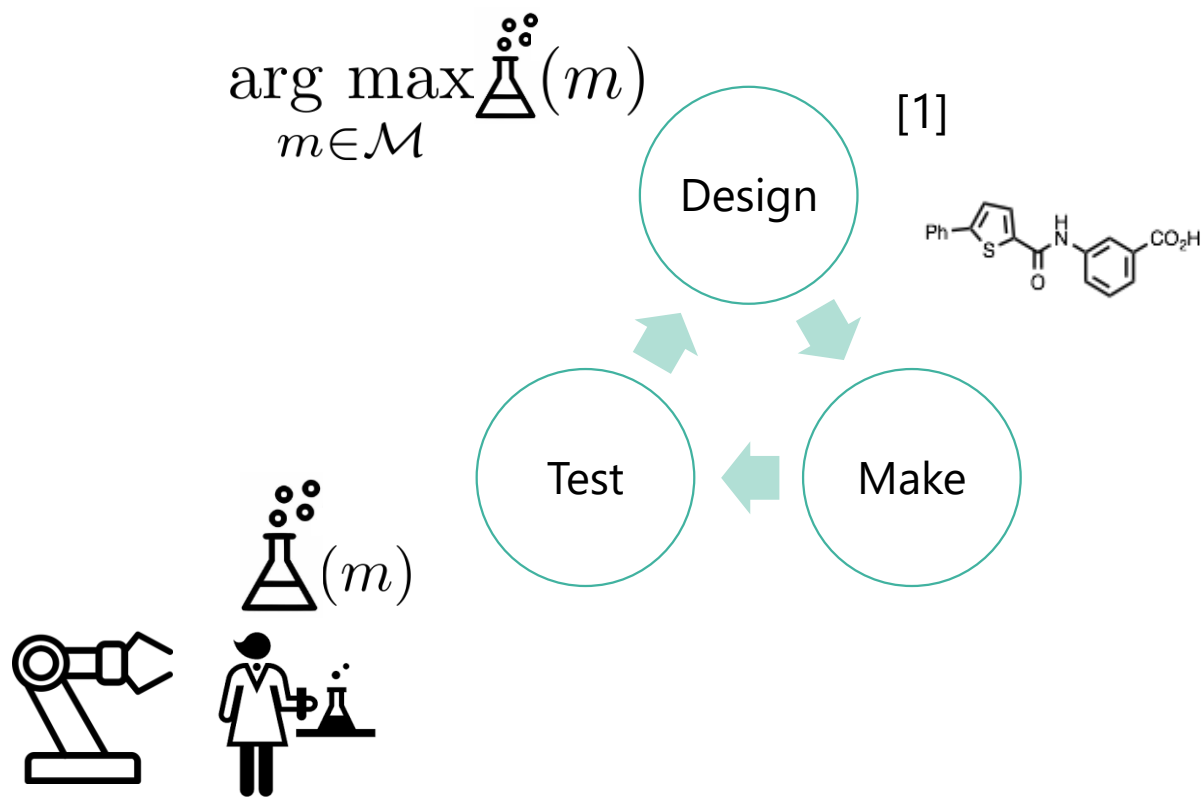
$$\arg \max_{m \in \mathcal{M}} \text{flask}(m)$$



Drug discovery, or why we need synthesis planning

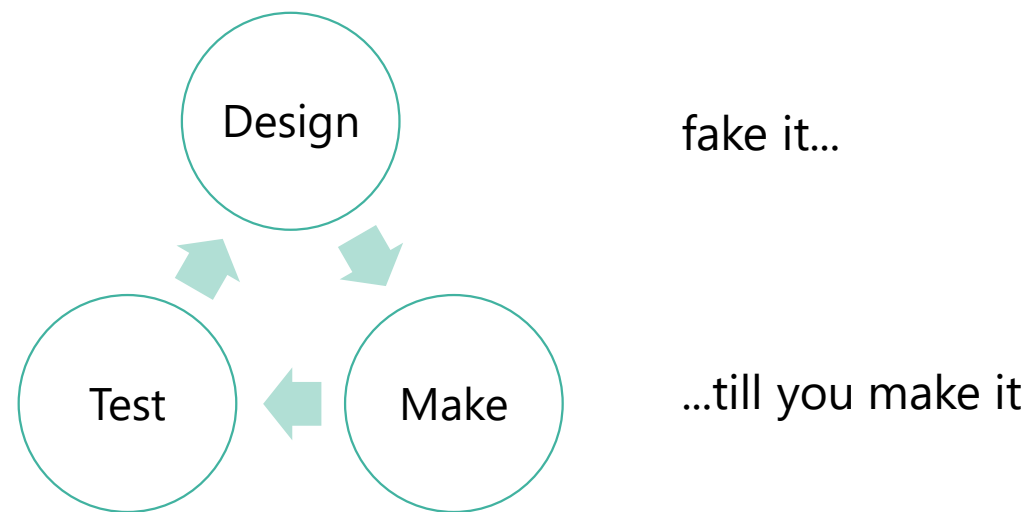


Drug discovery, or why we need synthesis planning

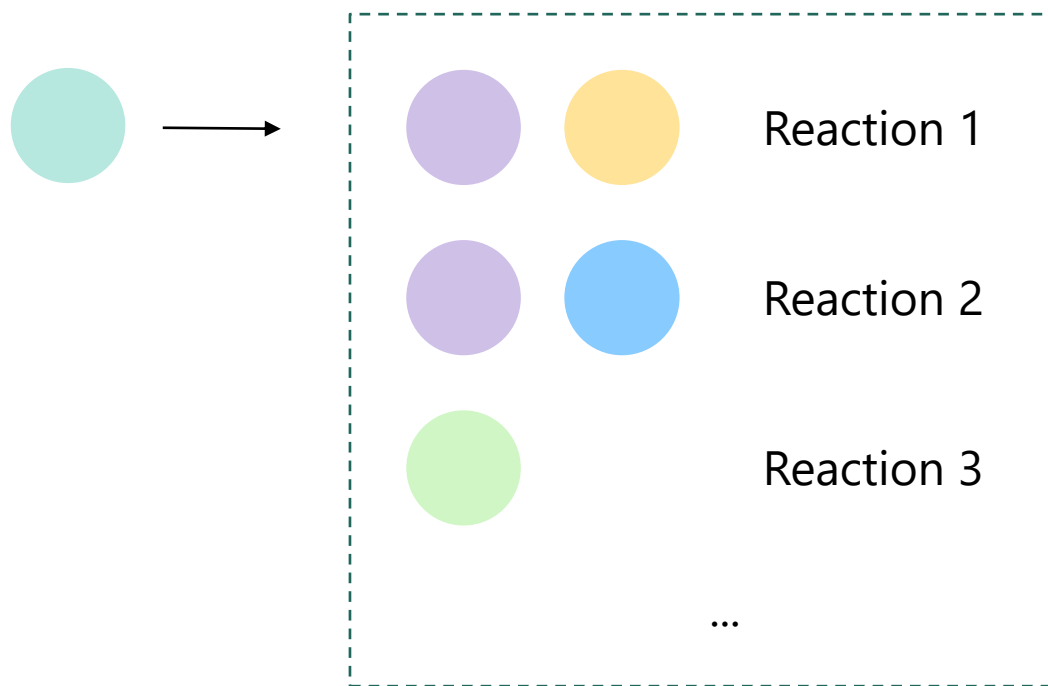


[1] Maziarz et al, "Learning to Extend Molecular Scaffolds with Structural Motifs", ICLR 2021

Drug discovery, or why we need synthesis planning



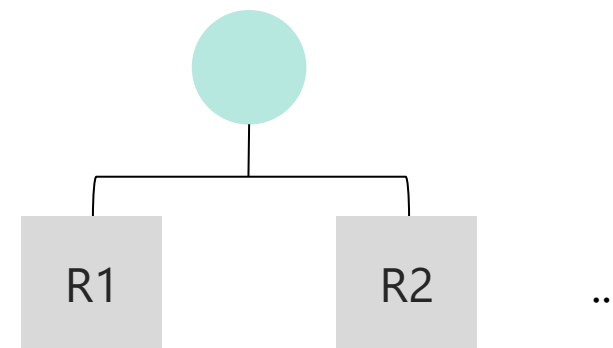
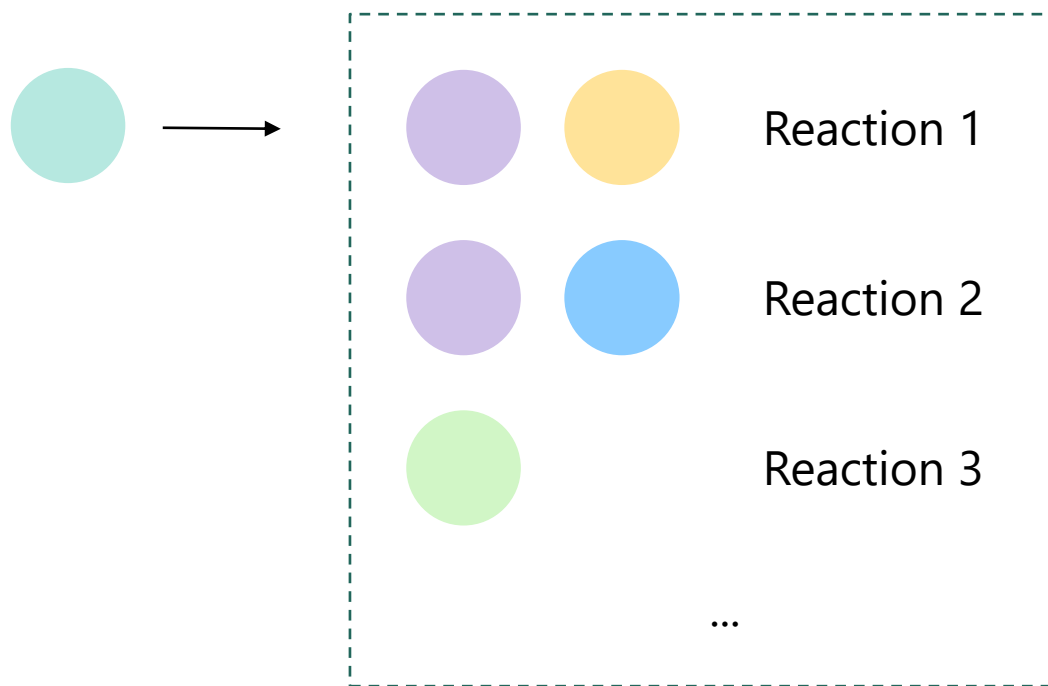
Retrosynthesis as a search problem



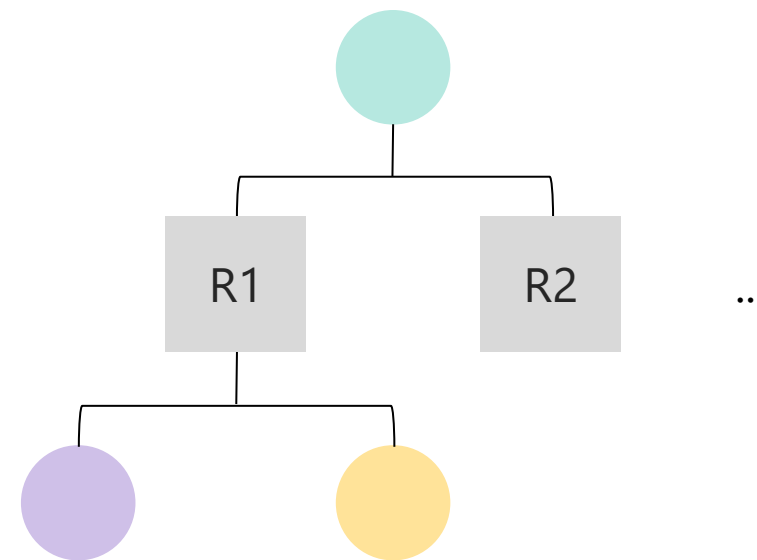
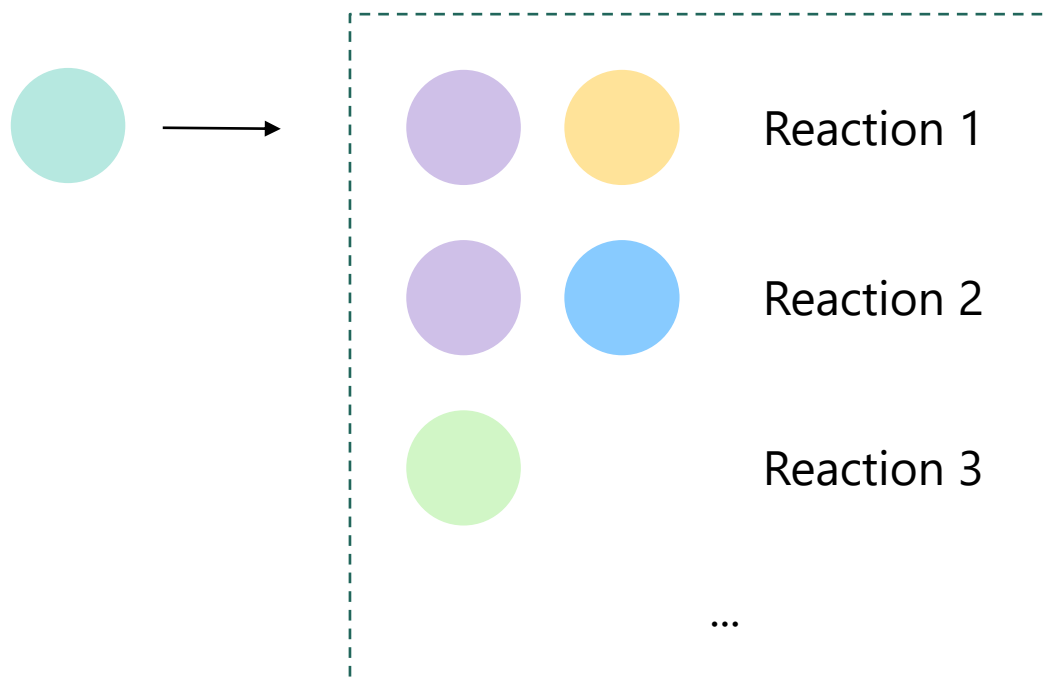
Retrosynthesis as a search problem



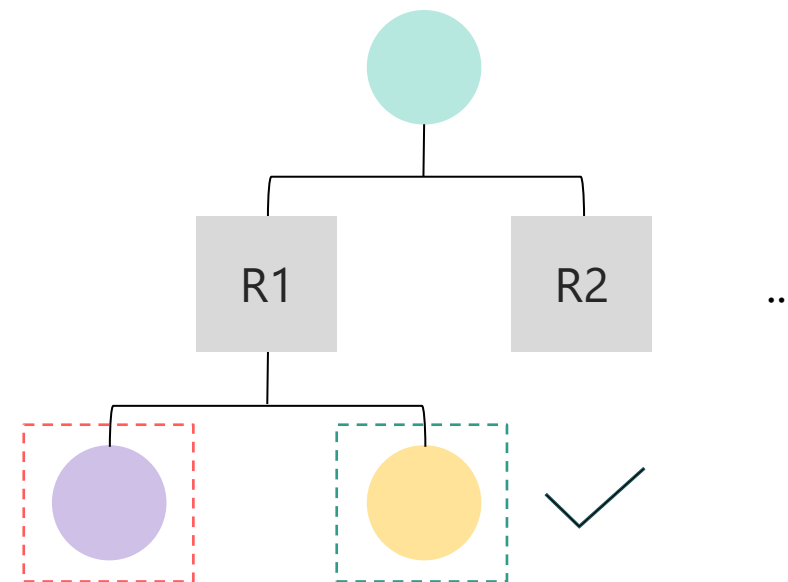
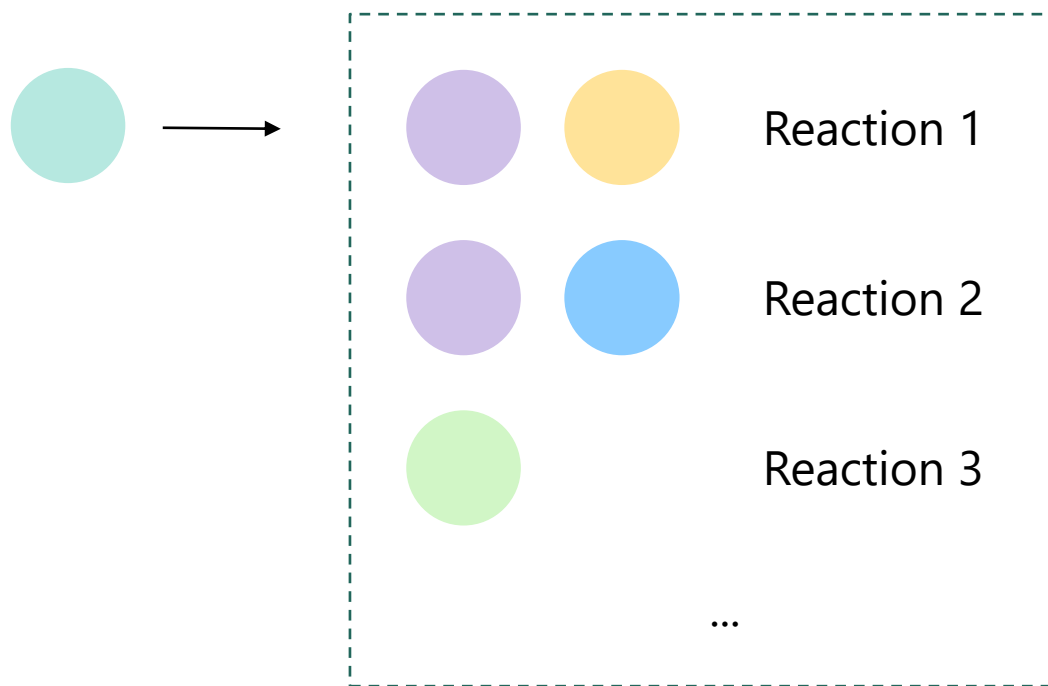
Retrosynthesis as a search problem



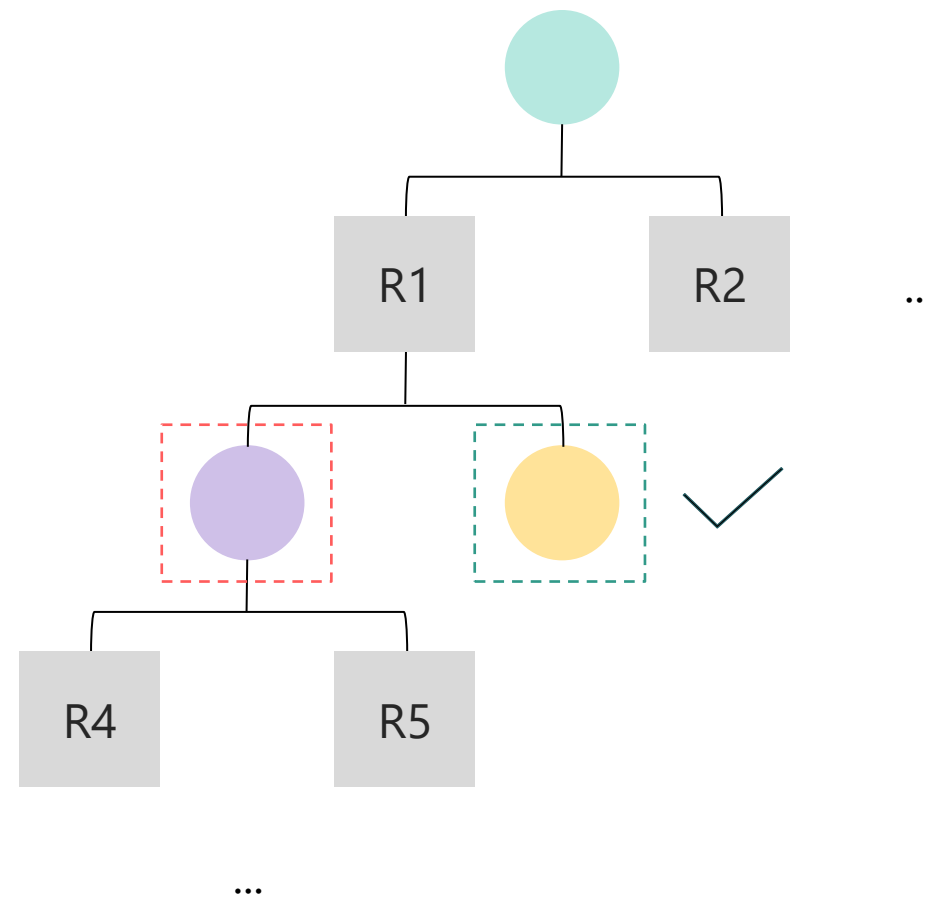
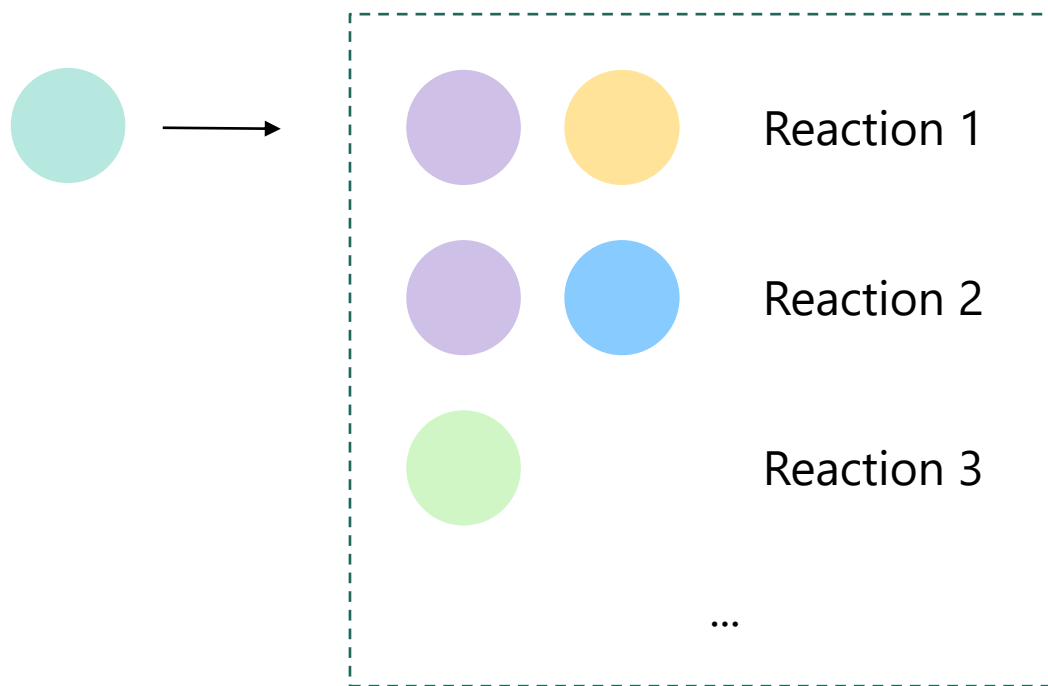
Retrosynthesis as a search problem



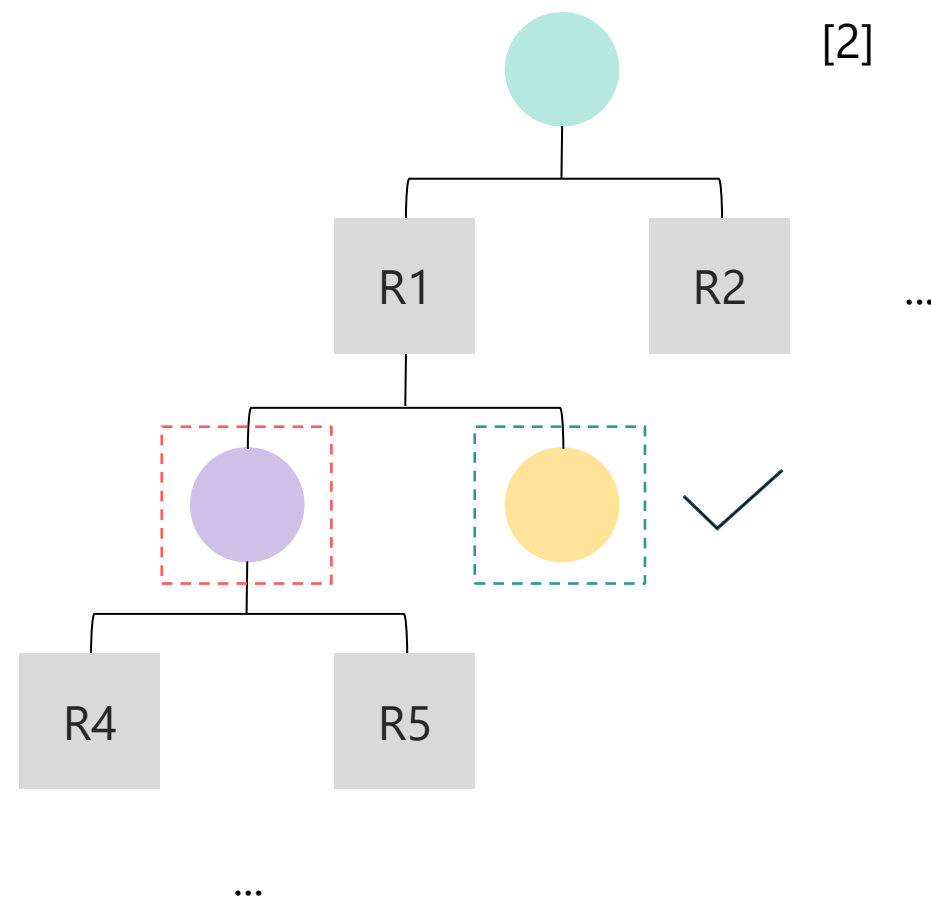
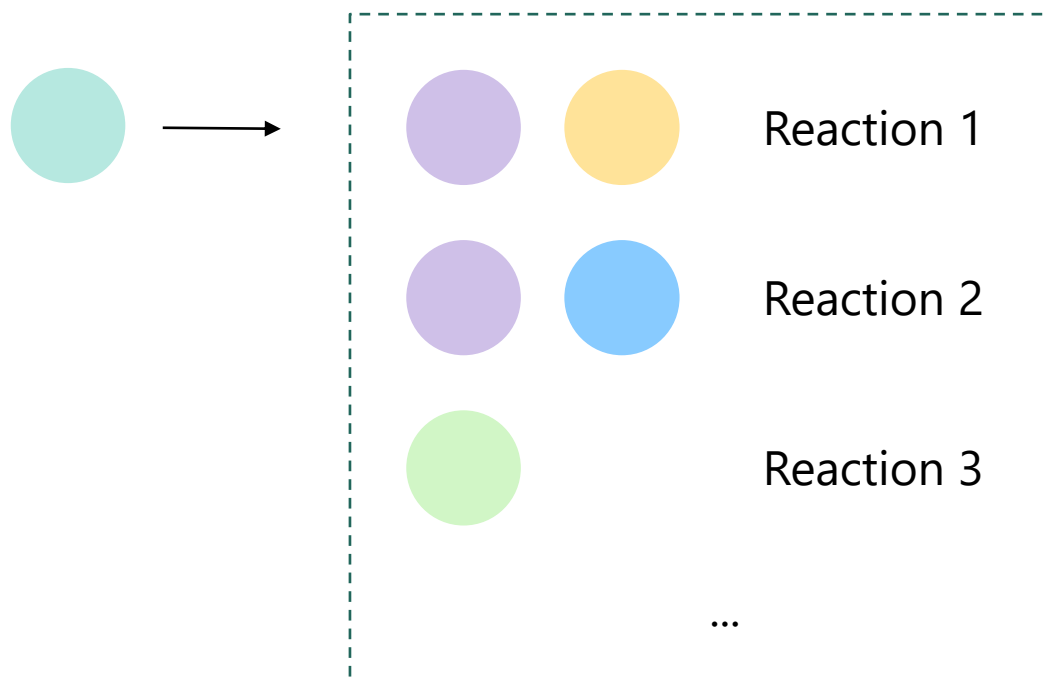
Retrosynthesis as a search problem



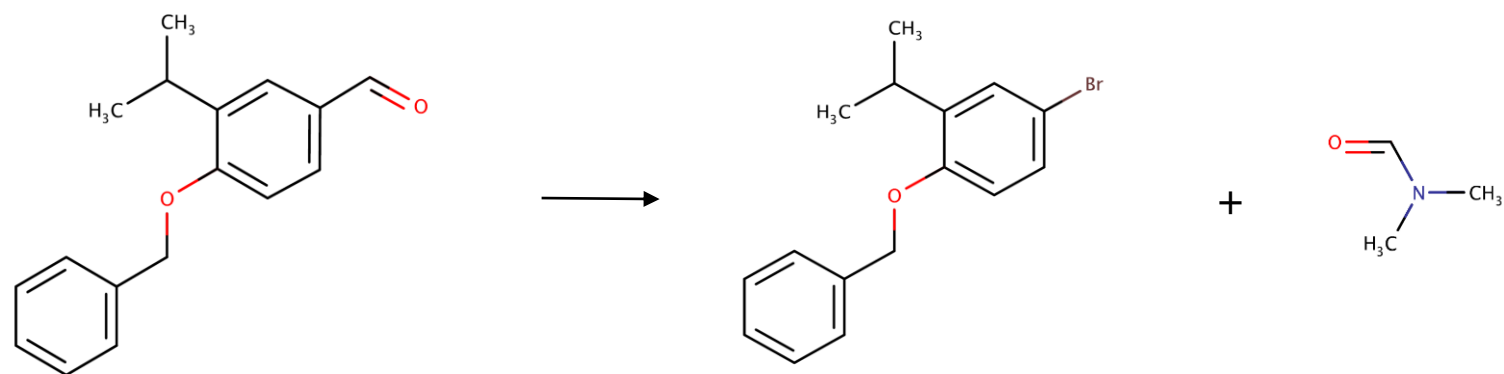
Retrosynthesis as a search problem



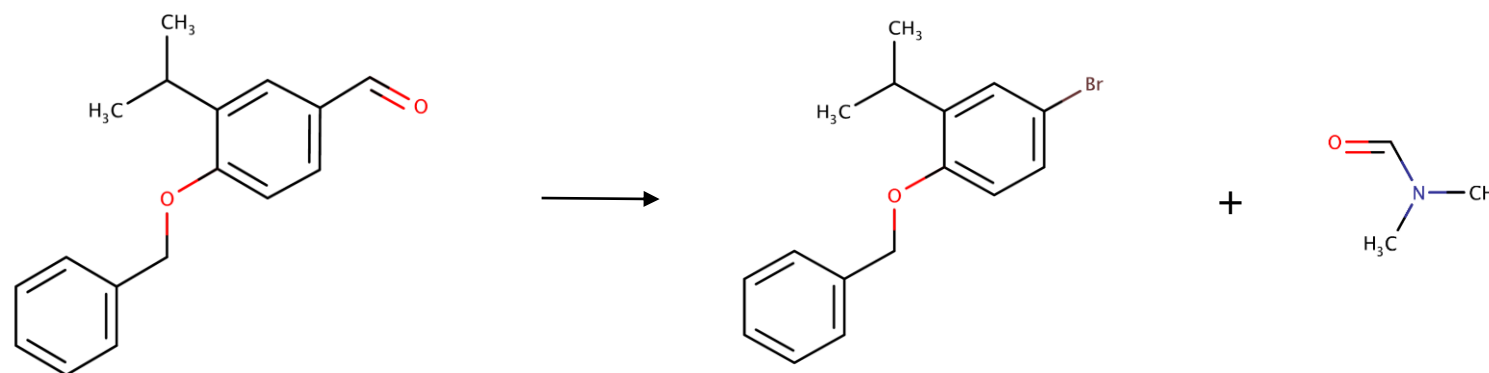
Retrosynthesis as a search problem



Single-step reaction prediction



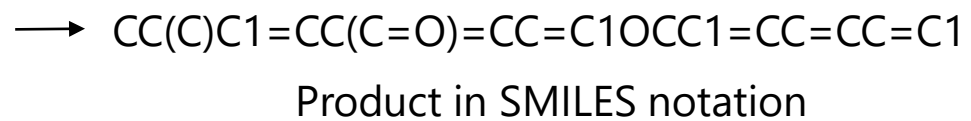
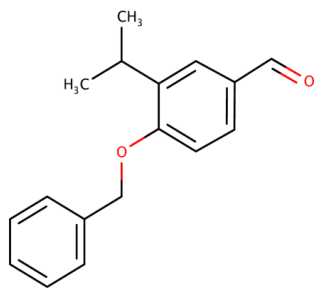
Single-step reaction prediction



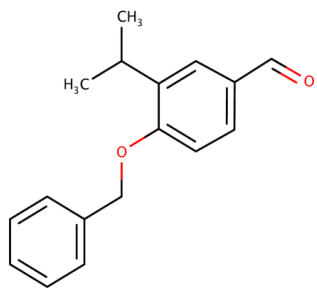
Main approaches:

- *end-to-end* → e.g. unconstrained Transformer
- *symbolic* → e.g. predicting the graph edit

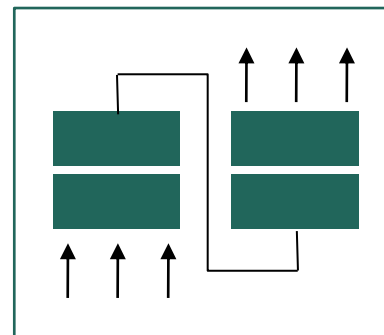
Single-step prediction: going end-to-end



Single-step prediction: going end-to-end

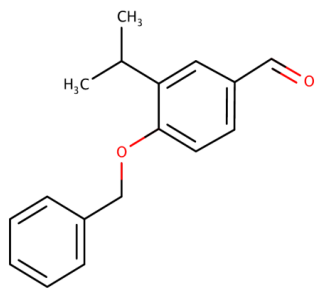


→ CC(C)C1=CC(C=O)=CC=C1OCC1=CC=CC=C1
Product in SMILES notation

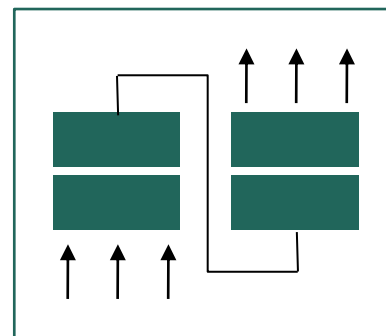


Encoder-Decoder Transformer

Single-step prediction: going end-to-end



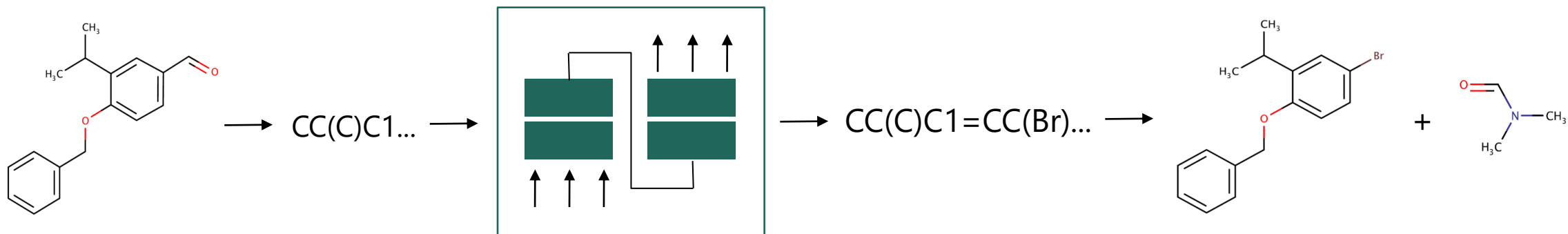
→ CC(C)C1=CC(C=O)=CC=C1OCC1=CC=CC=C1
Product in SMILES notation



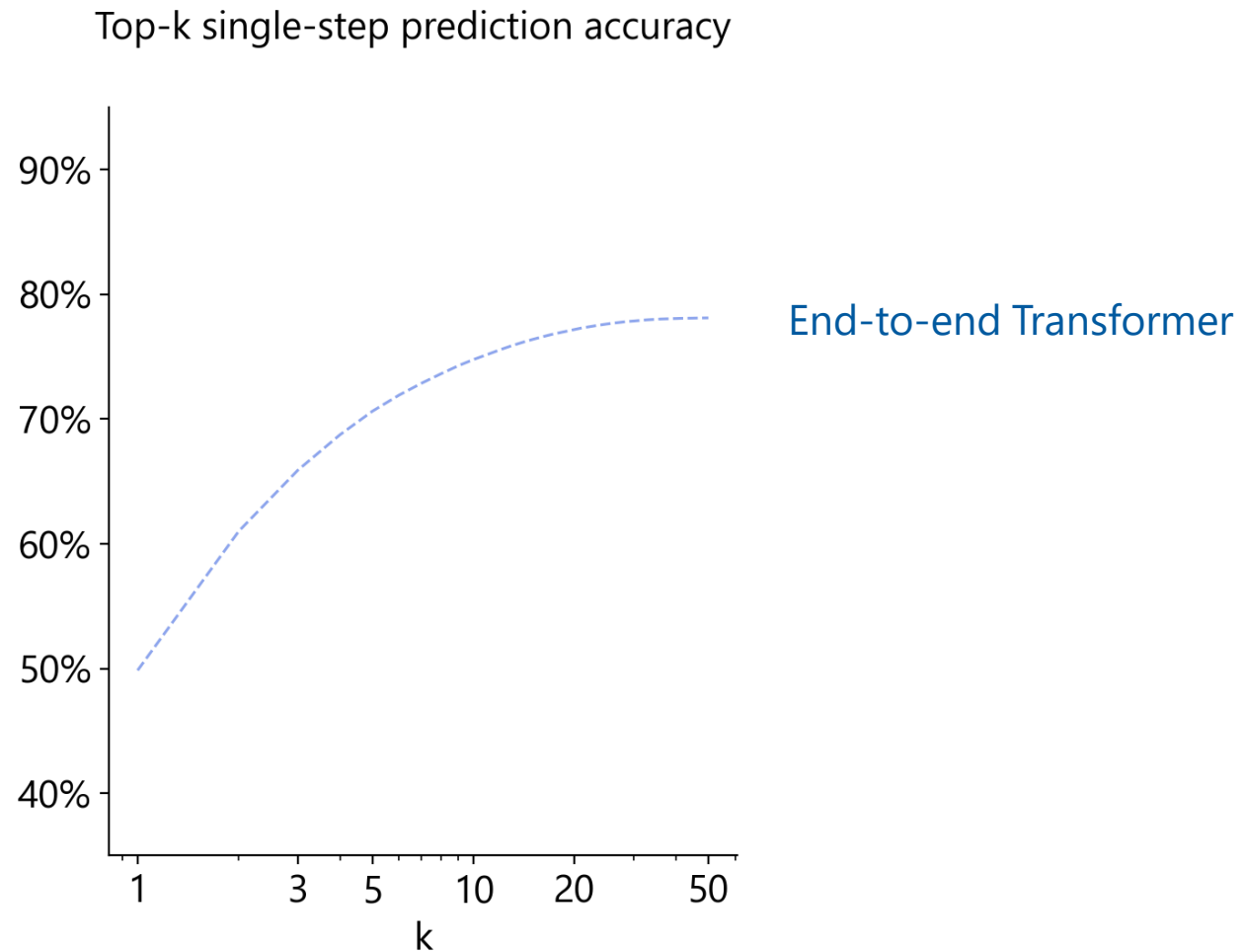
Encoder-Decoder Transformer

→ CC(C)C1=CC(Br)...

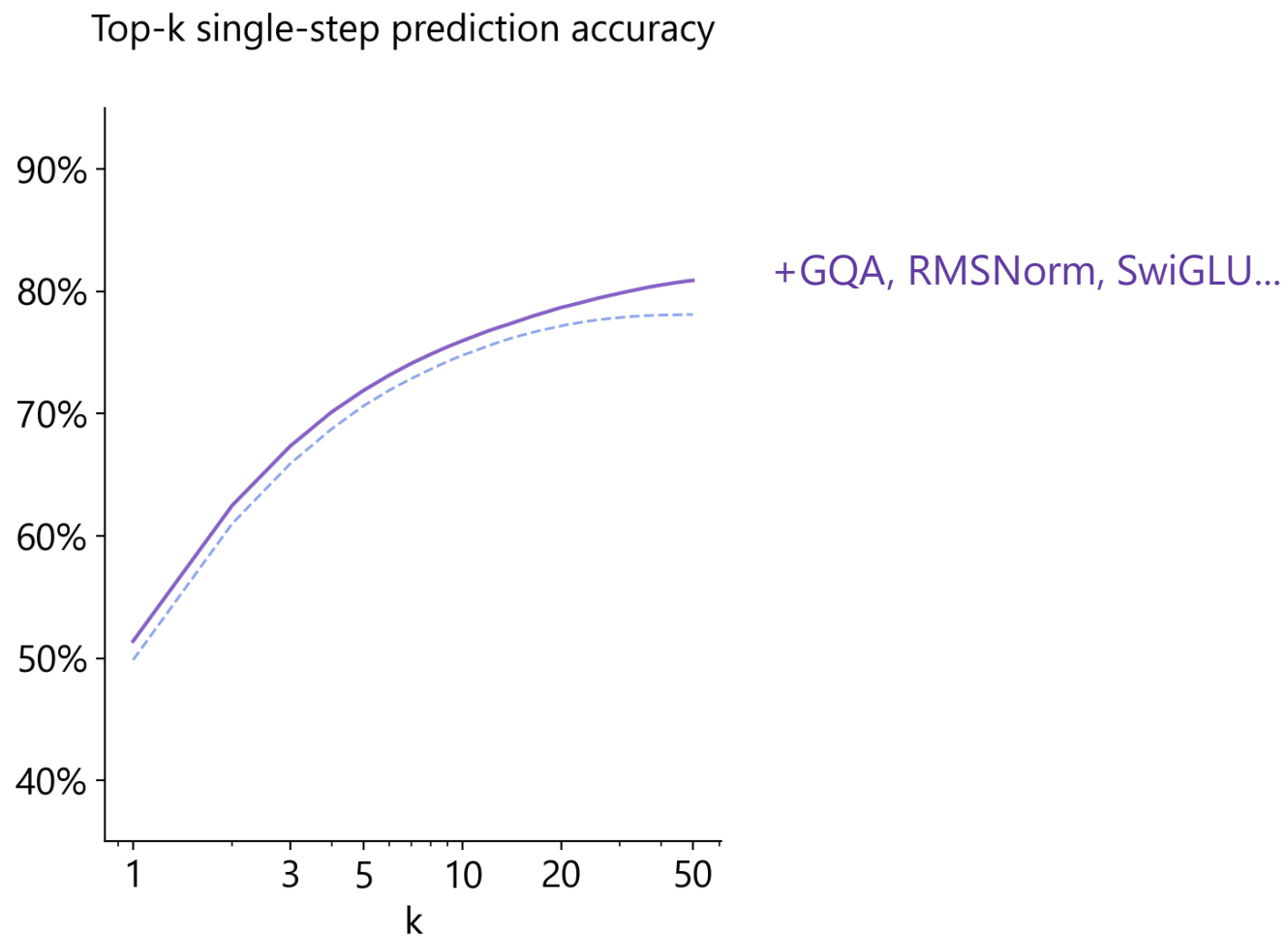
Single-step prediction: going end-to-end



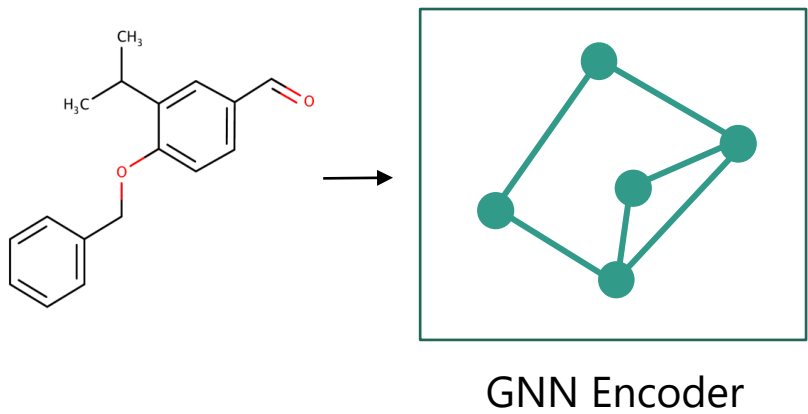
Single-step prediction: going end-to-end



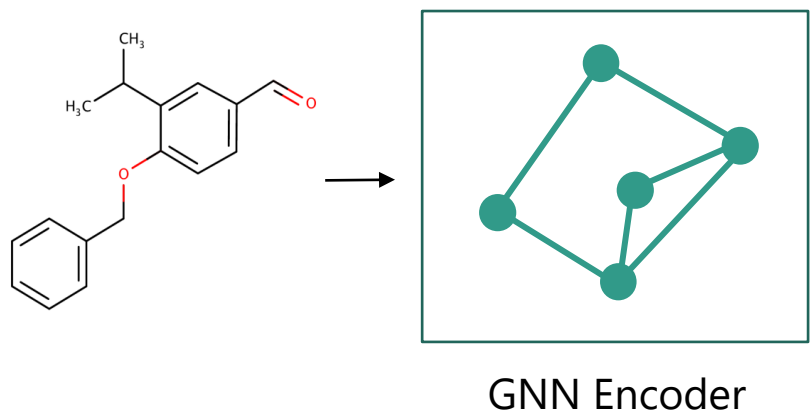
Single-step prediction: going end-to-end



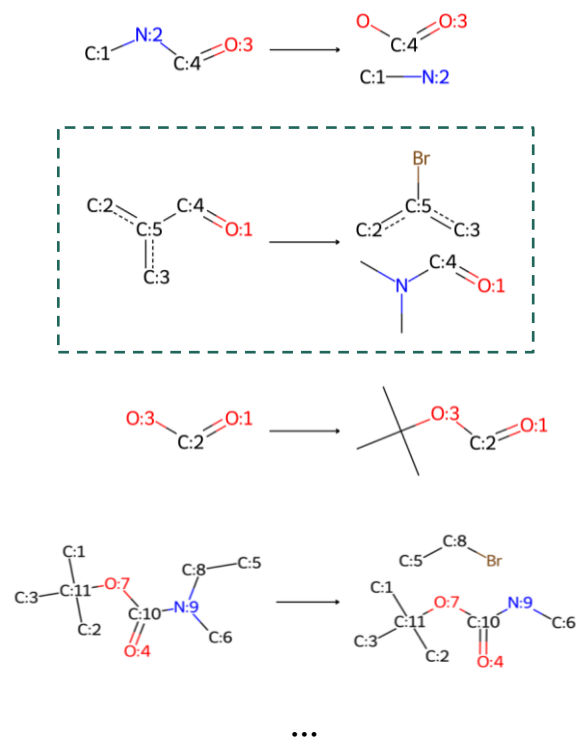
Single-step prediction: going symbolic



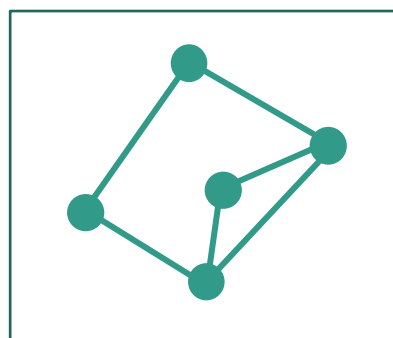
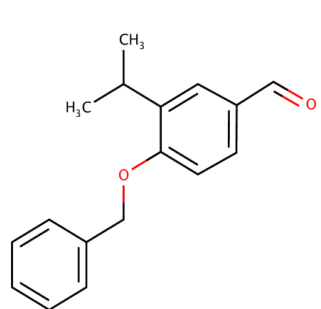
Single-step prediction: going symbolic



Template (graph rewrite) library



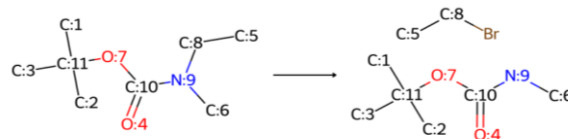
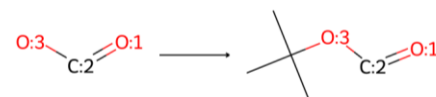
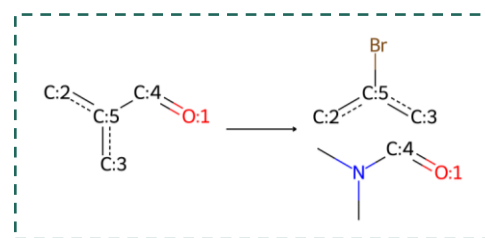
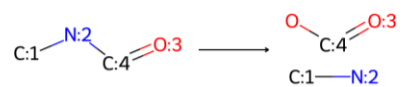
Single-step prediction: going symbolic



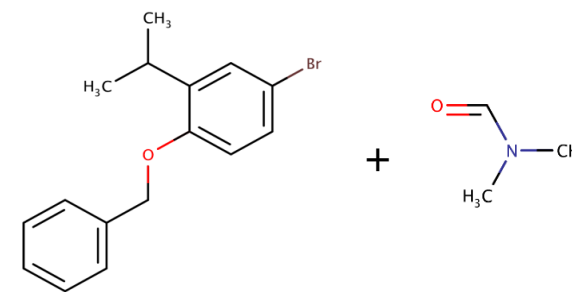
GNN Encoder



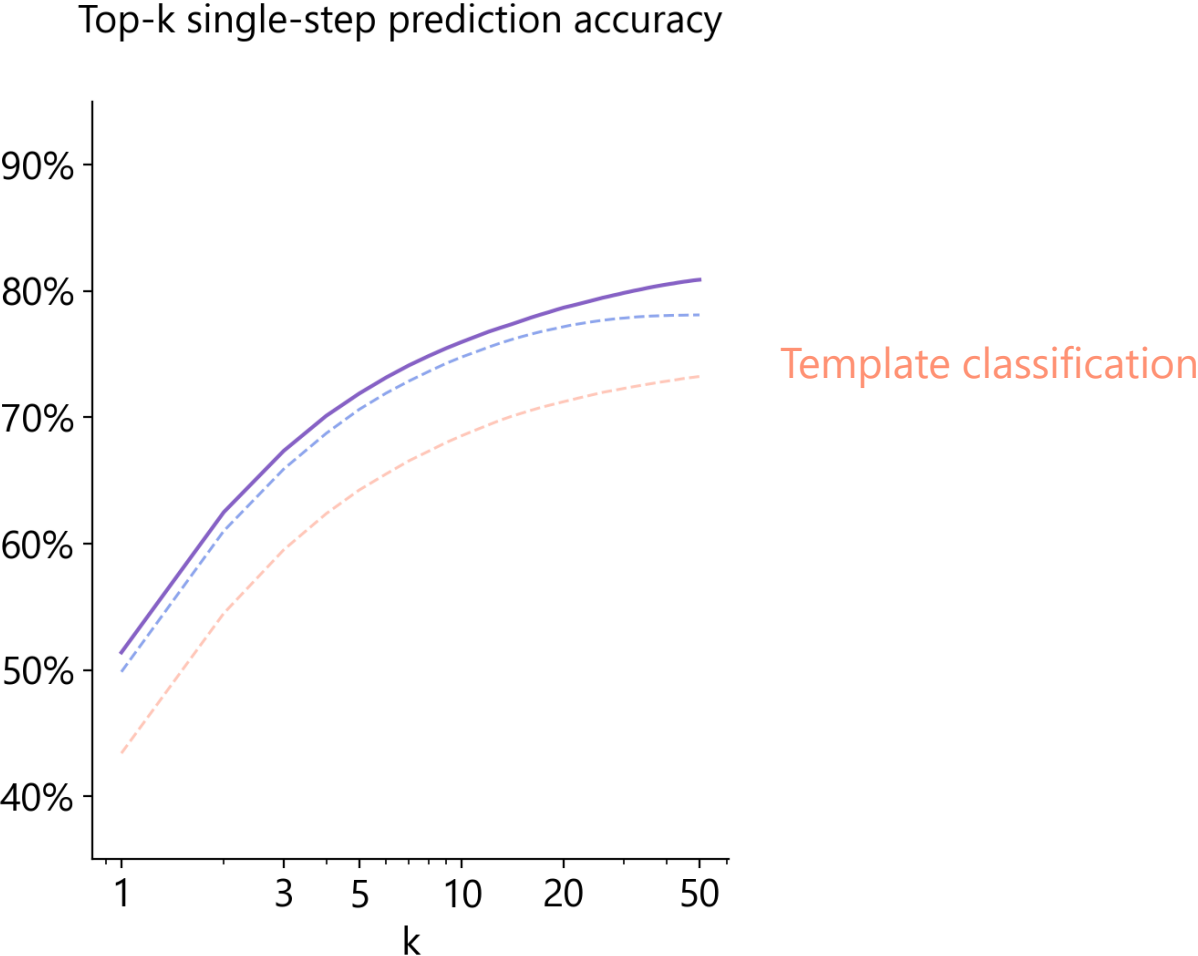
Template (graph rewrite) library



...

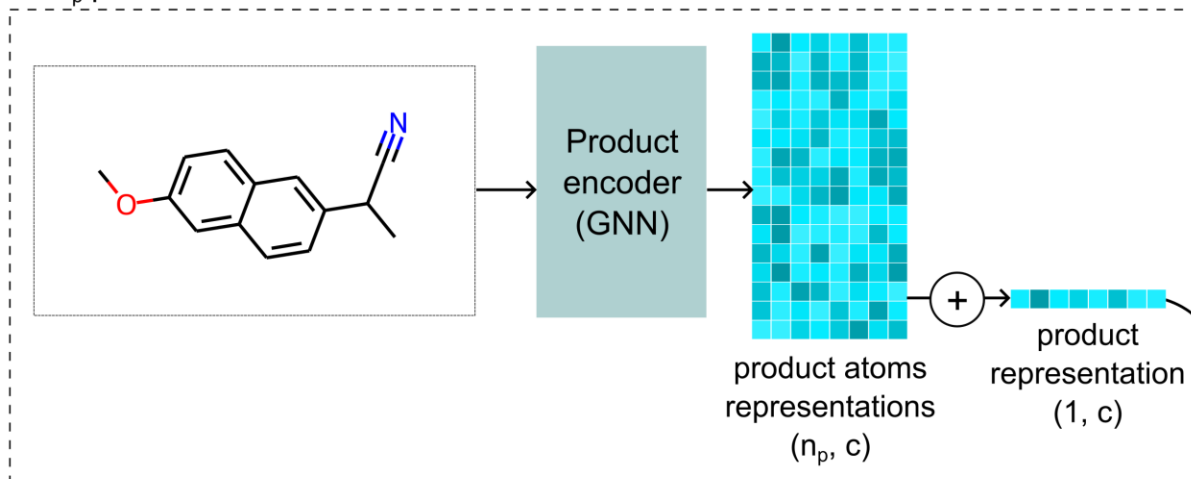


Single-step prediction: going symbolic

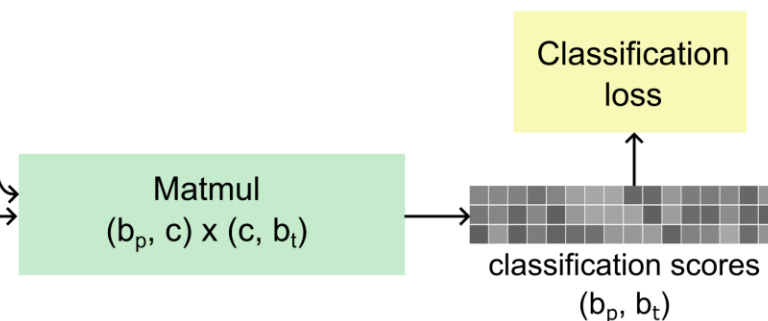
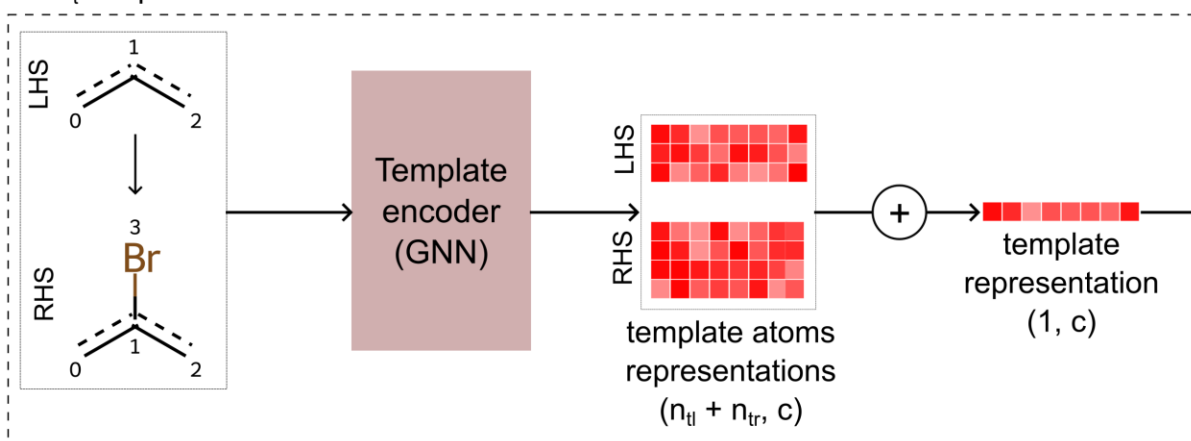


Single-step prediction: going symbolic

for b_p products in batch

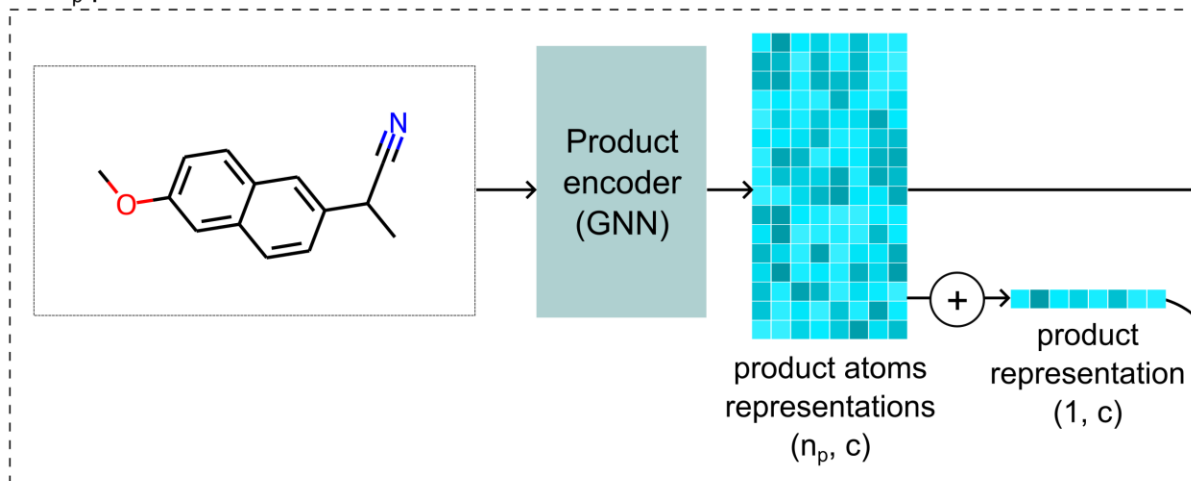


for b_t templates in batch

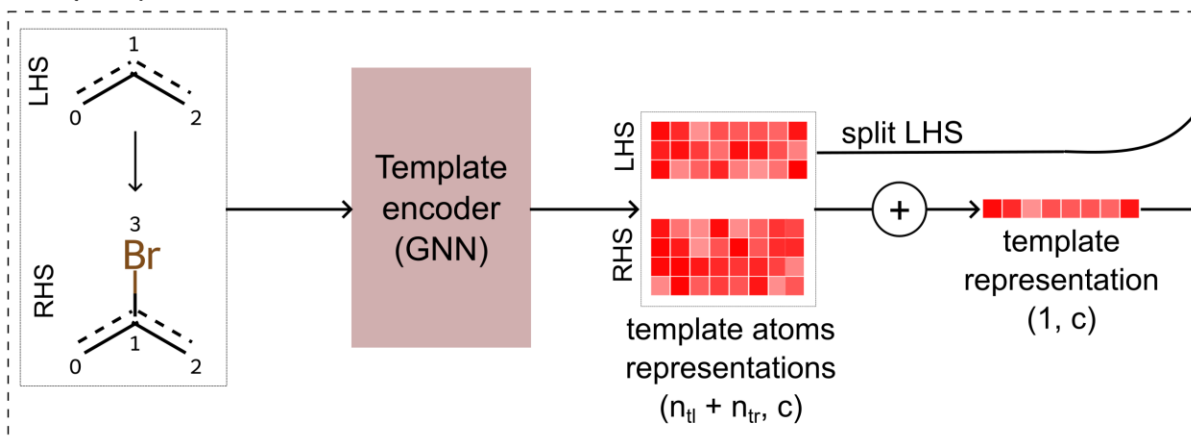


Single-step prediction: going symbolic

for b_p products in batch



for b_t templates in batch



Matmul (b_p times)
 $(n_p, c) \times (c, \sum n_{tl})$

localization scores
 $b_p \times (n_p, \sum n_{tl})$

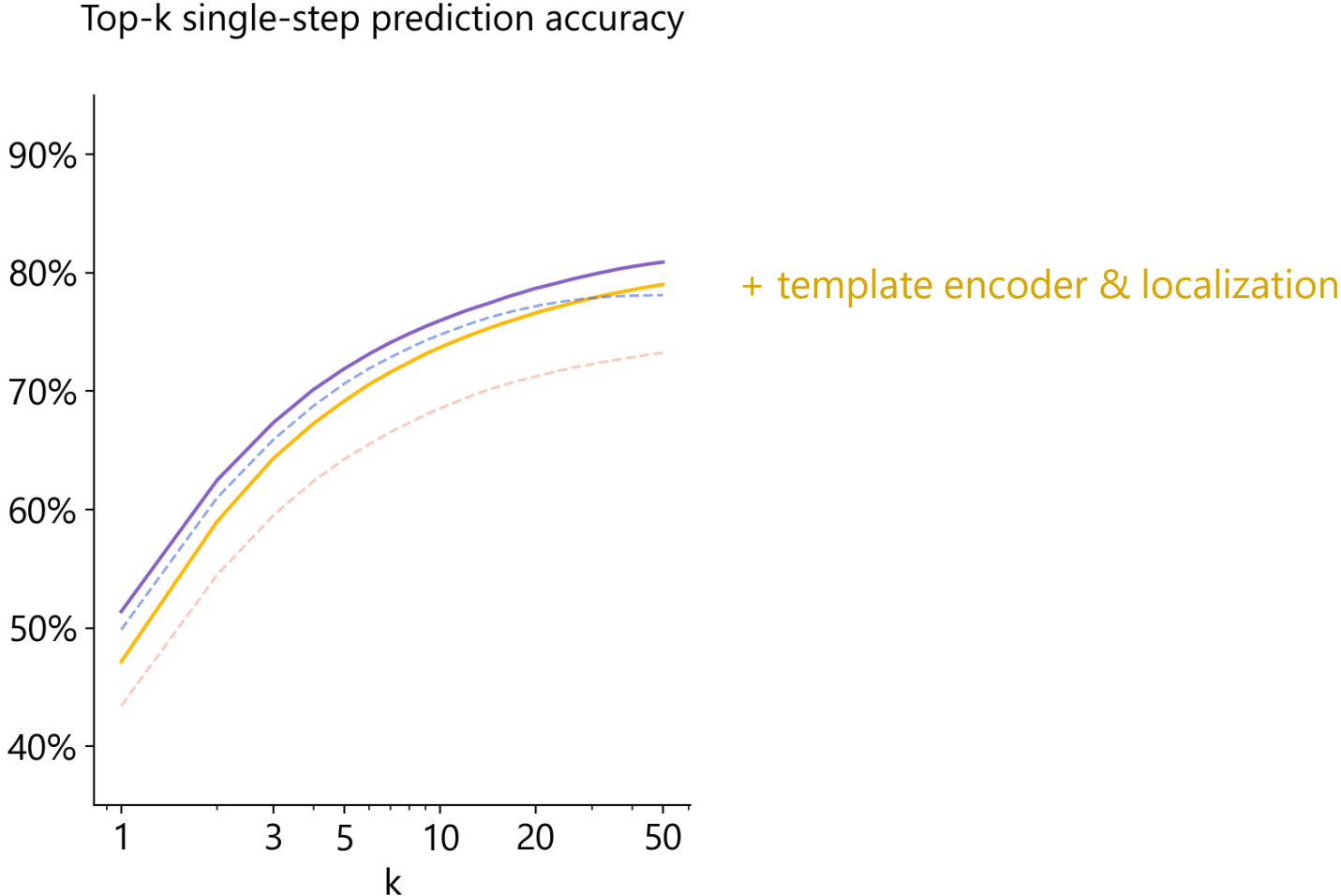
Localization loss

Matmul
 $(b_p, c) \times (c, b_t)$

classification scores
 (b_p, b_t)

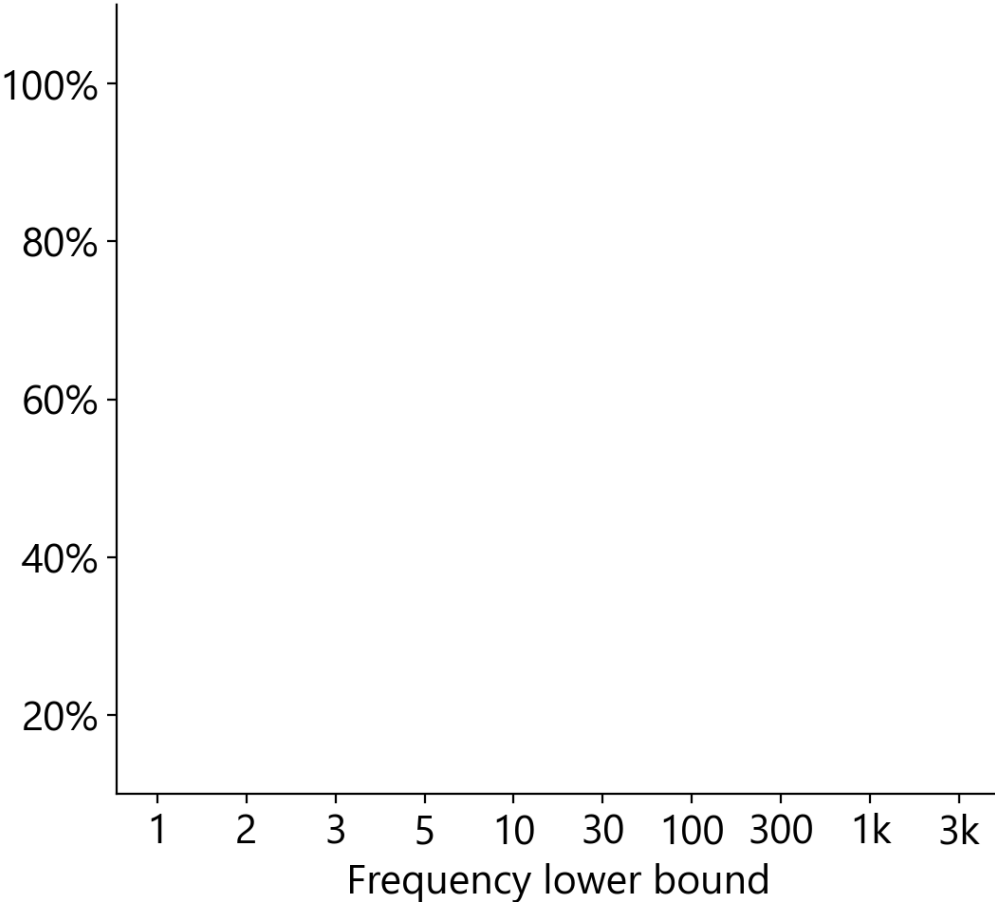
Classification loss

Single-step prediction: going symbolic



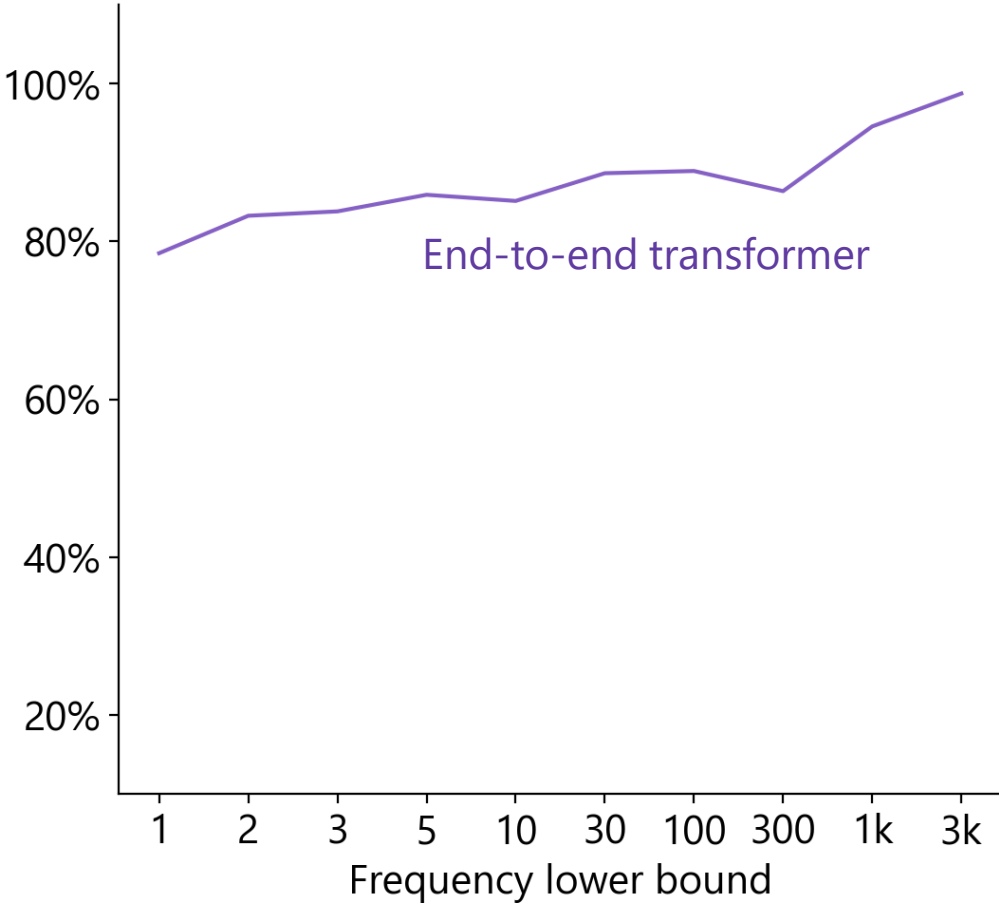
Hidden trade-offs

Top-50 accuracy vs template support



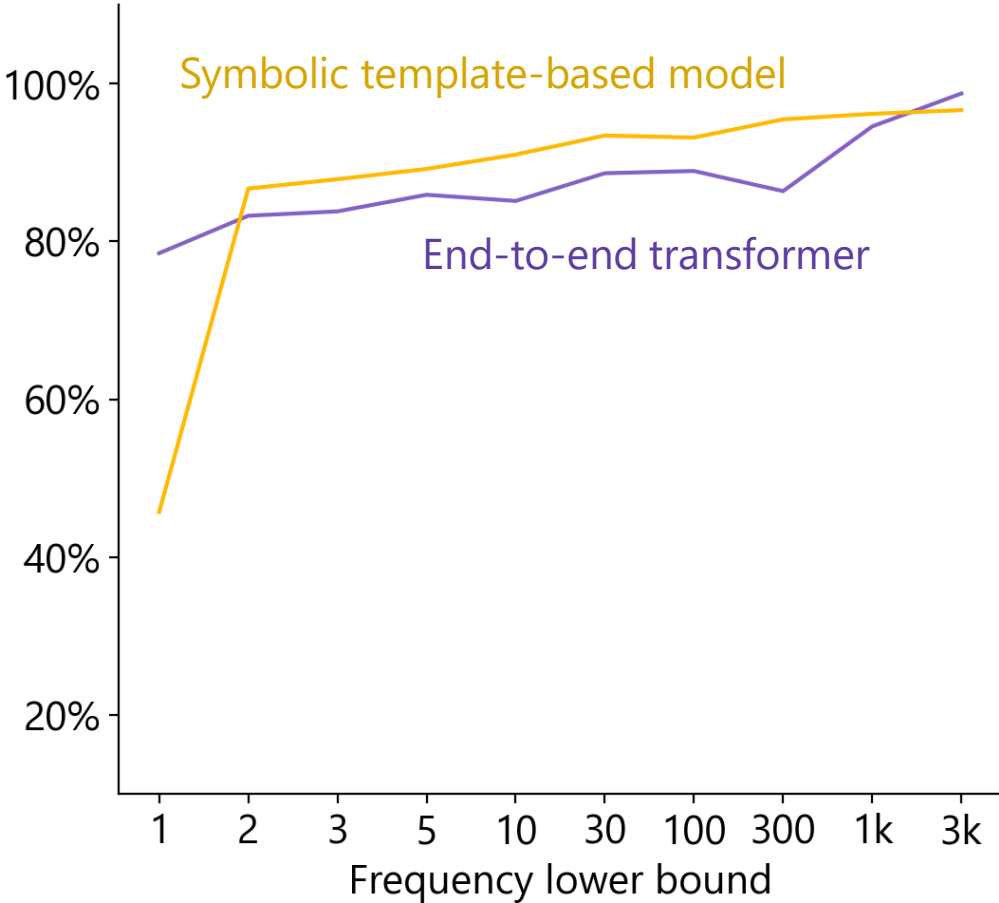
Hidden trade-offs

Top-50 accuracy vs template support

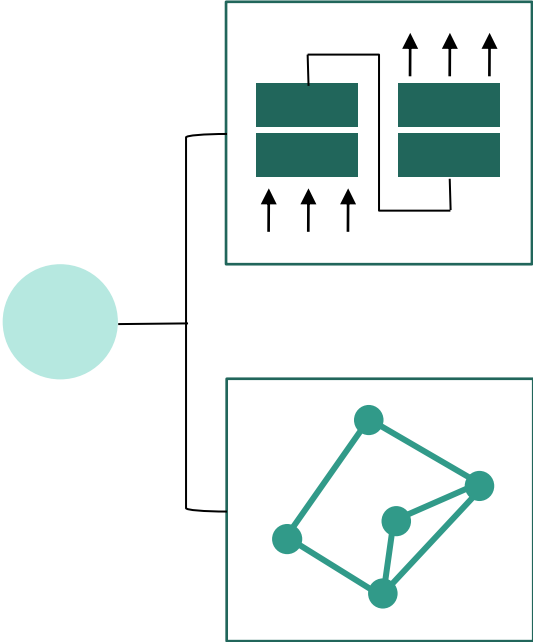


Hidden trade-offs

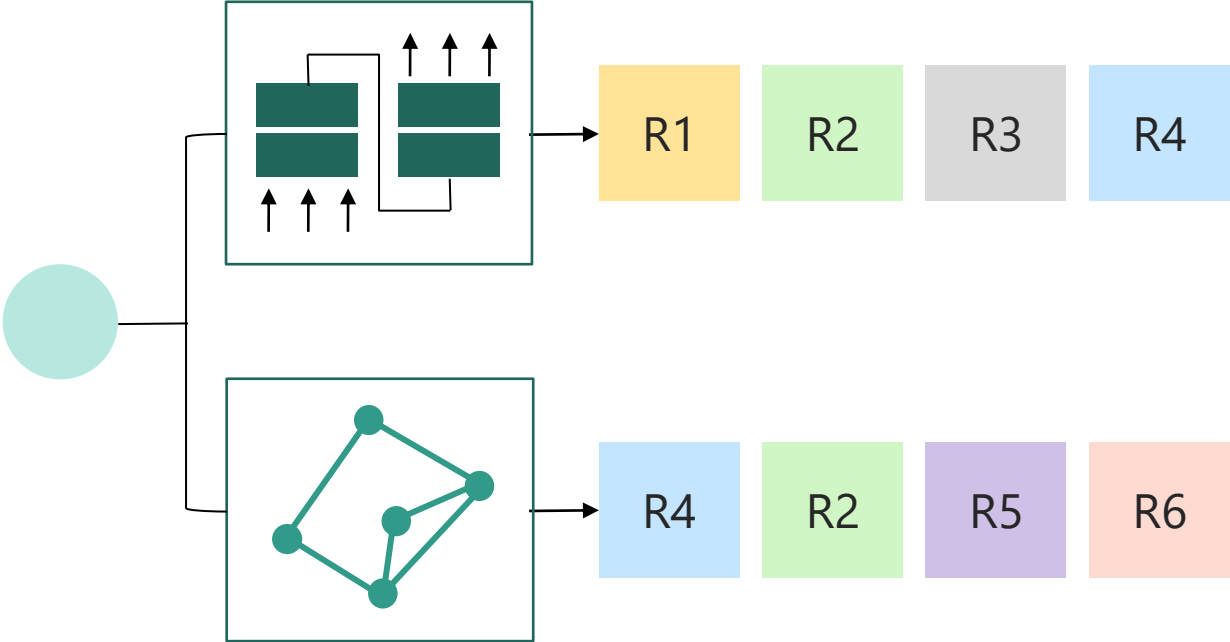
Top-50 accuracy vs template support



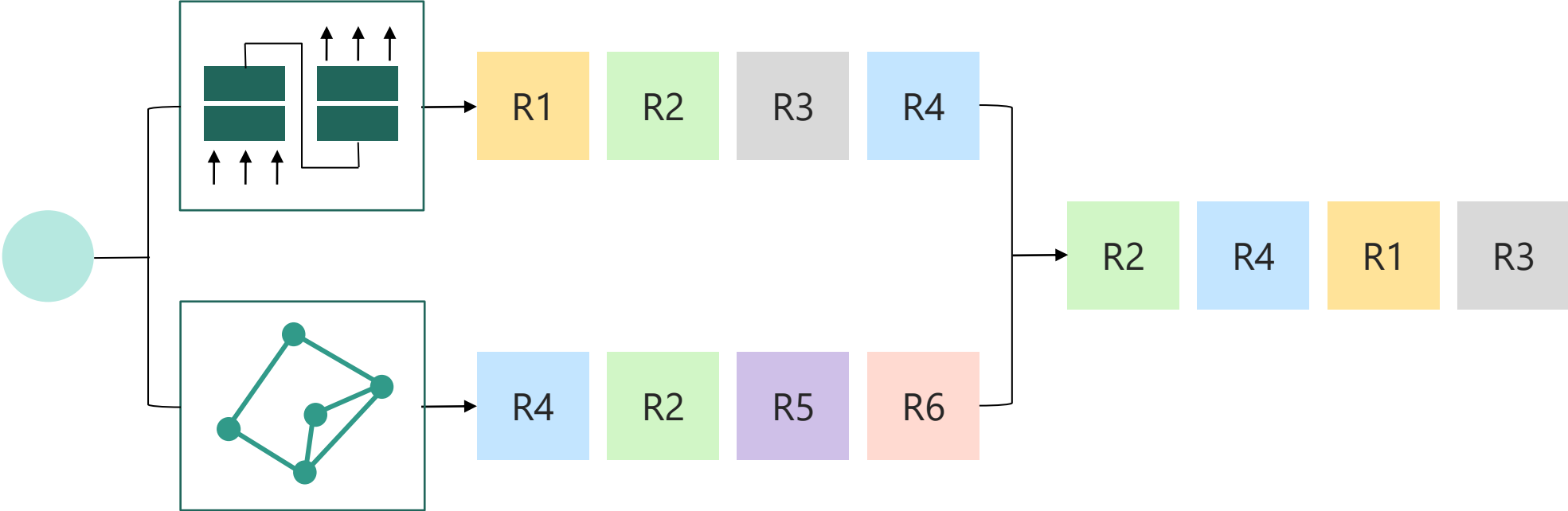
Chimera



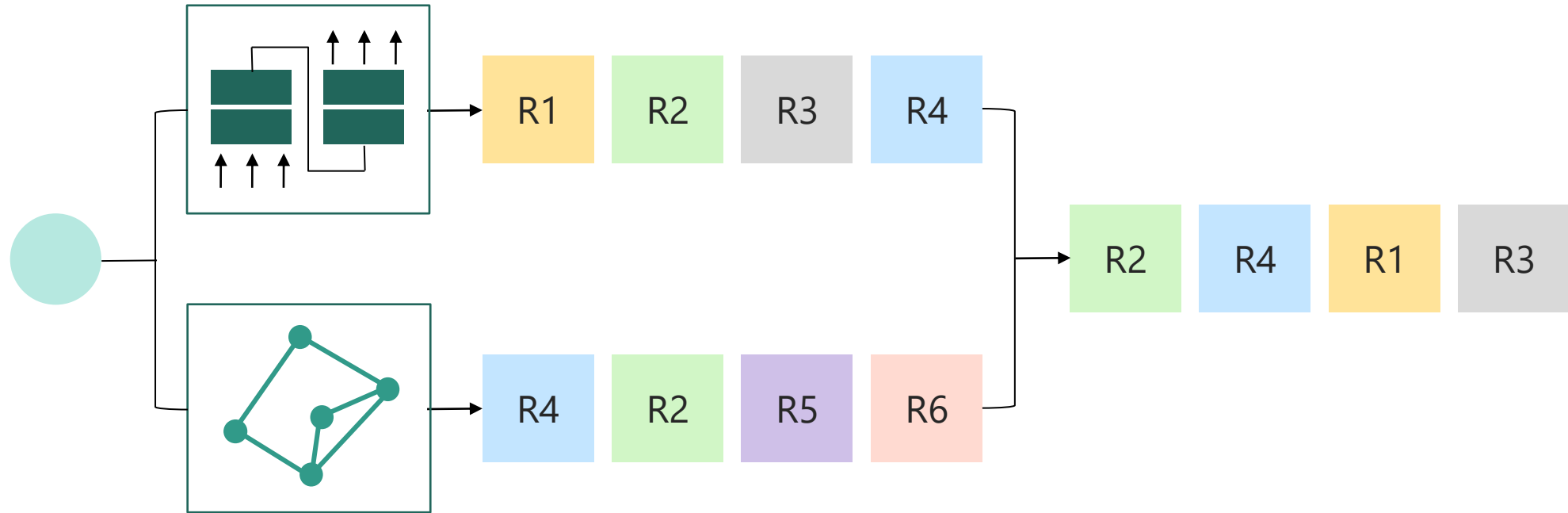
Chimera



Chimera



Chimera

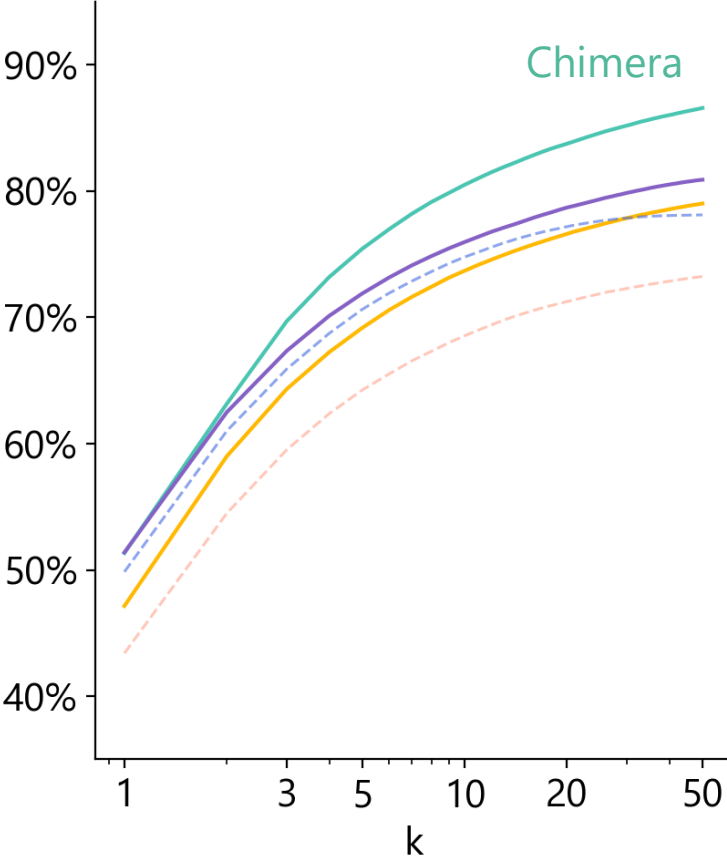


$$\text{score}(r) = \sum_{i=1}^m \sum_{k=1}^{k_{max}} \mathbb{1}[r = r_{i,k}] \cdot \theta_{i,k}$$

$$\mathcal{L}_{rank}(r^+, r^-) = \sigma \left(\frac{\text{score}(r^-) - \text{score}(r^+) + \epsilon}{T} \right)$$

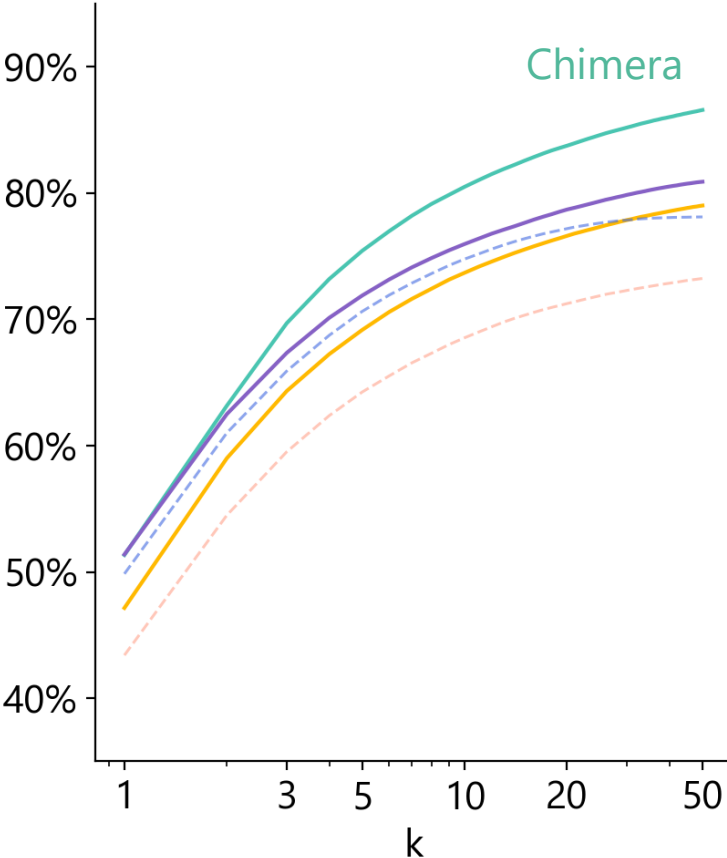
Chimera

Top-k single-step prediction accuracy

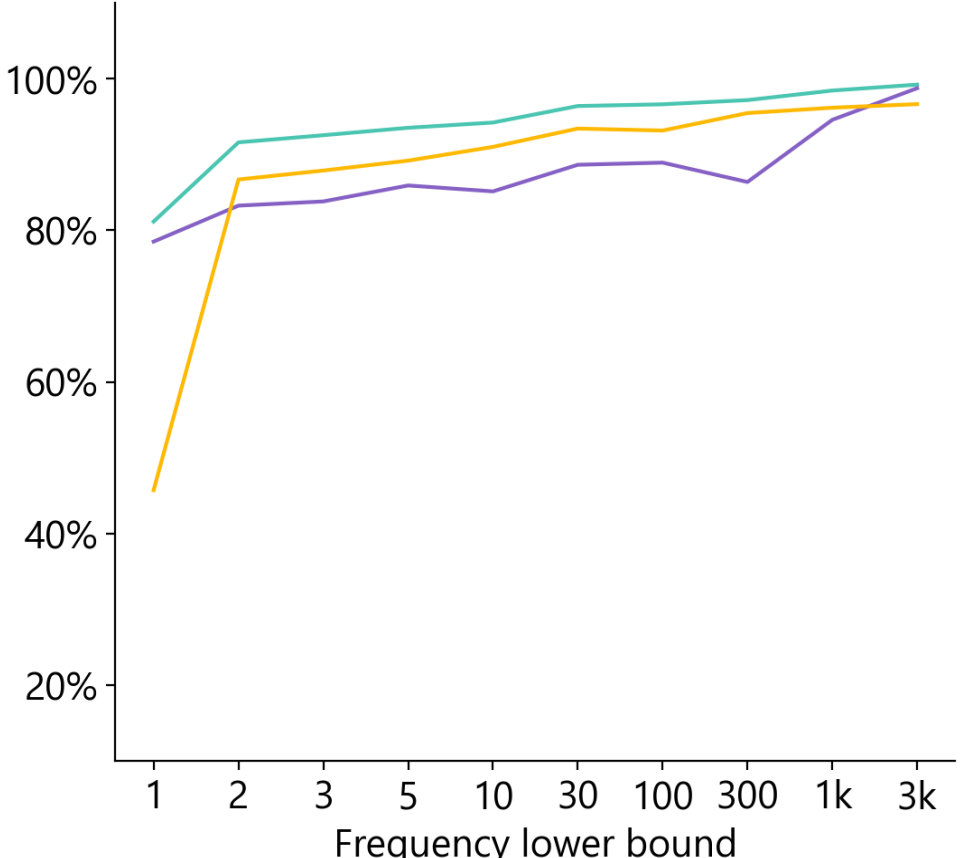


Chimera

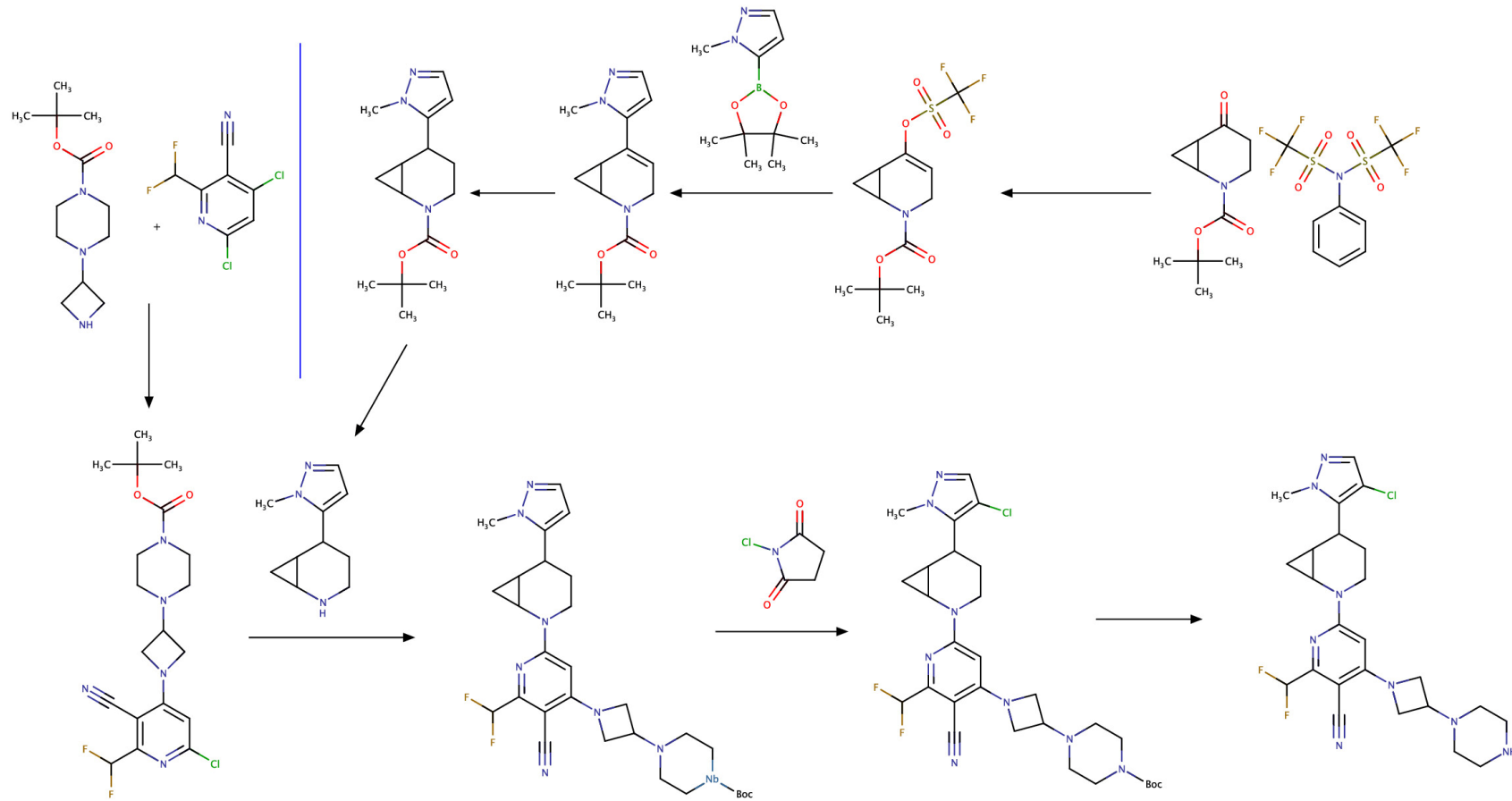
Top-k single-step prediction accuracy



Top-50 accuracy vs template support



Chimera



Acknowledgements:

Guoqing Liu (MSR)
Marwin Segler (MSR)
Hubert Misztela (Novartis)
Aleksei Kornev (Novartis)
Holger Hoeffling (Novartis)
Mike Fortunato (Novartis)
Rishi Gupta (Novartis)
Austin Tripp (U Cambridge)
Piotr Gaiński (Jagiellonian U)

We believe deep learning will have a transformational impact on the natural sciences

Questions

To learn more about Microsoft Research AI for Science, visit aka.ms/AI4Science



Amsterdam, Netherlands



Beijing, China



Berlin, Germany



Cambridge, UK



Redmond, USA

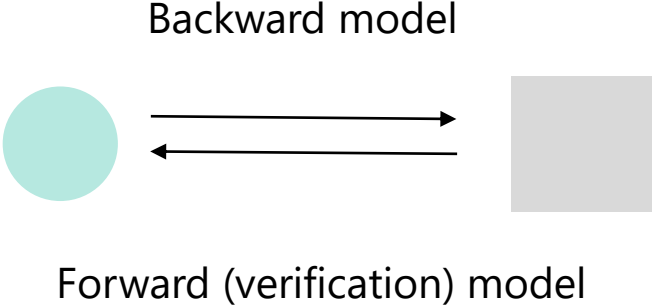


Shanghai, China

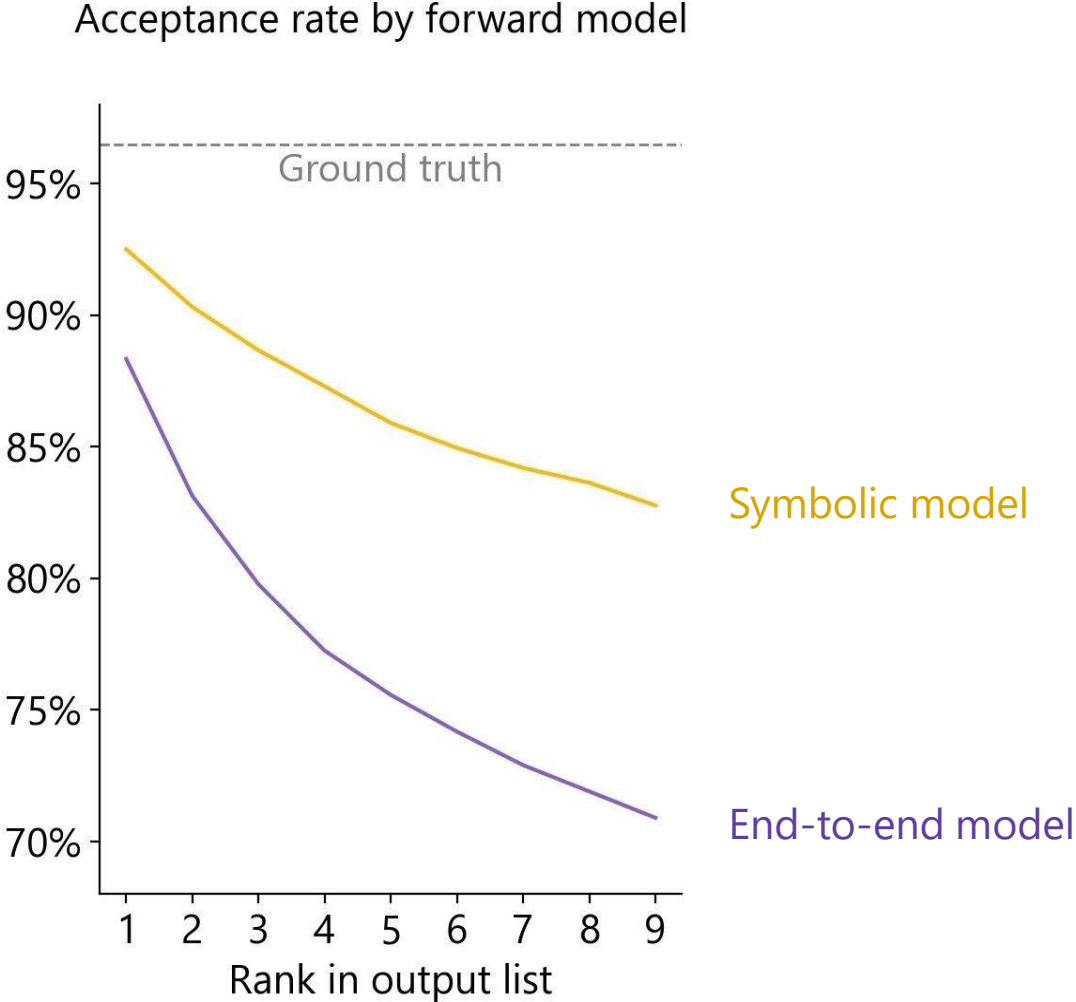
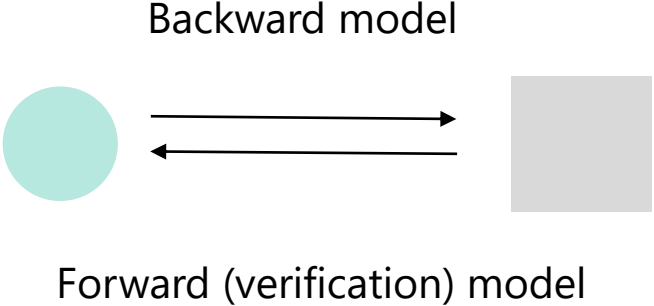


Backup slides

Beyond ground truth: verification via forward model



Beyond ground truth: verification via forward model



Beyond ground truth: verification via forward model

